

# delivering benefits through evidence



Making better use of local data in flood  
frequency estimation

Report – SC130009/R

We are the Environment Agency. We protect and improve the environment and make it a better place for people and wildlife.

We operate at the place where environmental change has its greatest impact on people's lives. We reduce the risks to people and properties from flooding; make sure there is enough water for people and wildlife; protect and improve air, land and water quality and apply the environmental standards within which industry can operate.

Acting to reduce climate change and helping people and wildlife adapt to its consequences are at the heart of all that we do.

We cannot do this alone. We work closely with a wide range of partners including government, business, local authorities, other agencies, civil society groups and the communities we serve.

This report is the result of research commissioned and funded by the joint Flood and Coastal Erosion Risk Management Research and Development Programme.

**Published by:**

Environment Agency, Horizon House, Deanery Road,  
Bristol, BS1 9AH  
[www.environment-agency.gov.uk](http://www.environment-agency.gov.uk)

ISBN: 978-1-84911-387-8

© Environment Agency – February 2017

All rights reserved. This document may be reproduced with prior permission of the Environment Agency.

Email: [fcerm.evidence@environment-agency.gov.uk](mailto:fcerm.evidence@environment-agency.gov.uk)

Further copies of this report are available from our publications catalogue:  
[www.gov.uk/government/publications](http://www.gov.uk/government/publications)

or our National Customer Contact Centre:  
T: 03708 506506

Email: [enquiries@environment-agency.gov.uk](mailto:enquiries@environment-agency.gov.uk)

**Authors:**

Harry Dixon, Duncan Faulkner, Matt Fry, Filip Kral, Rob Lamb, Mark Macklin, Ilaria Prosdocimi, Duncan Reed, Peter Rogers, Catherine Sefton, Lisa Stewart and Gianni Vesuviano

**Dissemination Status:**

Publicly available

**Research Contractor:**

JBA Consulting  
South Barn, Broughton Hall, Skipton, North Yorkshire  
BD23 3AE  
01756 799919

**Environment Agency's Project Manager:**

Sean Longfield and Chrissy Mitchell

**Theme Manager:**

Sue Manson, Incident Management and Modelling Theme

**Collaborators:**

Centre for Ecology & Hydrology  
DWRConsult  
University of Bath  
University of Aberystwyth

**Project Number:**

SC130009

# Evidence at the Environment Agency

Evidence underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us, helps us to develop tools and techniques to monitor and manage our environment as efficiently and effectively as possible. It also helps us to understand how the environment is changing and to identify what the future pressures may be.

Evidence is a key ingredient in the partnership between research, guidance and operations that enables the Environment Agency to protect and restore our environment by:

- **Setting the agenda**, by providing the evidence for decisions;
- **Maintaining scientific credibility**, by ensuring that our programmes and projects are fit for purpose and executed according to international standards;
- **Carrying out research**, either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available.

Doug Wilson

**Director of Research, Analysis and Evaluation**

# Executive summary

Flood frequency estimates are an essential part of flood risk management. Methods described in the Flood Estimation Handbook (FEH) published in 1999, and many subsequent updates, are the industry standard for flood frequency estimation in the UK.

Even carefully calculated flood frequency estimates are associated with many sources of uncertainty. These hydrological uncertainties often constitute the most uncertain component in any flood study. Uncertainty, where it is recognised, can lead to difficulty in having confidence in the outputs of studies such as flood map outlines, designs of flood defences or other structures, decisions on new development or information needed for investment planning or insurance. It can also lead to a loss of public credibility. As a result, there is considerable benefit to be gained from any reduction in the uncertainty of flood frequency estimation.

The raw material for flood estimation is high quality long-term records of river flow and rainfall. The FEH methods are based primarily on these types of data. There are many supplementary sources of information that can help to refine flood frequency and potentially reduce uncertainty. Examples include long-term flood history, river level records, temporary flow gauges, photographs of flood impacts, information obtained from field visits, measurements of channel width and evidence of flood deposits seen in the landscape (palaeoflood data). These and similar types of information are defined as local data. The FEH Local research project aimed to:

- quantify the uncertainty of design floods estimated from FEH methods
- develop procedures and guidance for incorporating local data into flood estimation to reduce such uncertainties

This report describes the review of scientific developments and good practice, and the development of new procedures carried out during the FEH Local project. A companion output from the project is a document, 'Using Local Data to Reduce Uncertainty in Flood Frequency Estimation', giving guidance to practitioners on how to estimate uncertainty in flood frequency and how to find and incorporate local data.

Following a wide-ranging review of the availability and use of local data, this report focuses on the evaluation and development of procedures for incorporating 2 main types of local data:

- historical (and palaeoflood) information
- channel dimensions

A statistical simulation study examines methods for flood frequency analysis using historical data and tests the sensitivity of the results to uncertainty of aspects such as the length of the historical period or the possibility of missing some events. The study recommends a maximum likelihood technique for combining historical and gauged flood data. The technique is able to incorporate either historical floods for which discharges can be estimated or floods for which all that is known is that the discharge exceeded a given threshold. It can also be used with palaeoflood data and this report includes a critical review of the use of palaeoflood data in UK practice.

A procedure is presented for estimating the median annual flood using a combination of catchment descriptors and bankfull channel width. This is an extension of a technique presented in the FEH using the same dataset of channel widths.



The practitioner guidance explains how to implement these procedures and provides case studies showing how historical data, palaeofloods and several other types of local data can be incorporated in flood studies.

The report presents a proposal for a new system to improve access to local data to be integrated with the National River Flow Archive. The development of this system is feasible, given clearly defined limits on the data types to be included and secure funding, with a national remit, both for the establishment and for the long-term operation and maintenance of the system.

Also covered in this report is a separate aspect of the FEH Local project, a pilot study to develop high-resolution catchment descriptors and explore the potential for new catchment descriptors to replace or augment some of those currently used in FEH methods. Catchment boundaries and descriptors are derived from a 10m digital terrain model, considerably more detailed than the 50m terrain data used in the FEH. Existing FEH catchment descriptors are evaluated for all catchments down to a minimum size of 0.2km<sup>2</sup> for a pilot area in north Cumbria. Several new descriptors are proposed, although it is not yet clear that they would necessarily lead to improved flood estimates.

There is now a challenge for the flood management sector to put into practice the findings of FEH Local, so that it becomes common practice to seek and exploit local data (rather than, as currently, best practice if it is done at all). This will lead to several benefits:

- better estimates of design flows
- reduced uncertainty
- project results that are more robust to challenge
- less need to seek reviews and improvement of hydrology studies
- enhanced public credibility

Ultimately, it can be expected that an outcome will be improved protection of people and property.

There are several ways to help meet this challenge, including dissemination of the practitioner guidance and encouragement or requirement of its implementation via appropriate wording of project scoping documents by the Environment Agency and other regulators. Appropriate use of local data should also be required in sector-specific guidance or specifications.

More rigorous reviews of flood estimates will help to challenge poor practice. There is a need for a change in culture that gets hydrologists out from their computer models more often and into the field.

# Acknowledgements

The project team is grateful to many individuals across the spectrum of academic to practical hydrology for inspiring ideas, fruitful discussions and constructive feedback.

Particular thanks are extended to:

- the Environment Agency's project board: Chrissy Mitchell, Tim Hunt, Julie Stannett and Sean Longfield
- those who have reviewed outputs: Glenda Tudor-Ward (Natural Resources Wales), Alistair Cargill (SEPA), David Worth (Royal Haskoning), David MacDonald (independent), Andy Wallis (Black and Veatch), Jamie Cooper (Staffordshire County Council) and Peter Spencer, Louise Glover, John Phillips, Steve Moore and Duncan Wishart (all Environment Agency)

Thomas Kjeldsen (University of Bath), a member of the project team, contributed valuable expertise on the evaluation of uncertainty. Geraldene Wharton (Queen Mary, University of London) supplied data on channel width and ideas on its application, and Paul Brewer (Aberystwyth University) and Anna Jones (Edge Hill University) provided palaeoflood data. Andrew Black, Frank Law and Celia Kirby contributed ideas on archiving flood history data, and David Morris contributed to the work on high-resolution catchment descriptors.

Those who have made more substantial contributions to the research itself are listed as authors.

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b>  |
| <b>2</b> | <b>Uncertainty in design flood estimation</b>                           | <b>2</b>  |
| 2.1      | Protocols for managing uncertainty                                      | 2         |
| 2.2      | Applications of uncertainty assessment in the flood management industry | 4         |
| 2.3      | Estimating uncertainty  | 5         |
| 2.4      | Uncertainty in the FEH statistical method                               | 6         |
| 2.5      | Uncertainty in the ReFH2 method   | 19        |
| <b>3</b> | <b>Review of local data for flood frequency estimation</b>              | <b>25</b> |
| 3.1      | Introduction  | 25        |
| 3.2      | Definition of local data  | 25        |
| 3.3      | Types of local data   | 25        |
| 3.4      | Perspectives on the use of local data                                   | 26        |
| 3.5      | UK guidance on the use of local data                                    | 28        |
| 3.6      | River flow and water levels   | 28        |
| 3.7      | Example: The Devon Hydrology Strategy                                   | 33        |
| 3.8      | Climate   | 34        |
| 3.9      | Groundwater levels  | 36        |
| 3.10     | Recent floods   | 37        |
| 3.11     | Historical floods   | 38        |
| 3.12     | Palaeofloods  | 41        |
| 3.13     | Catchment properties  | 42        |
| 3.14     | Channel dimensions  | 44        |
| 3.15     | Examples of complex flood behaviour                                     | 47        |
| <b>4</b> | <b>Development of procedures for enhanced use of local data</b>         | <b>49</b> |
| 4.1      | Summary of local data types   | 49        |
| 4.2      | Statistical procedures for incorporating historical or palaeoflood data | 51        |
| 4.3      | Estimating QMED from channel dimensions                                 | 55        |
| 4.4      | Proposal for improving access to local data                             | 71        |
| <b>5</b> | <b>High-resolution catchment descriptors</b>                            | <b>74</b> |
| 5.1      | Introduction  | 74        |
| 5.2      | Drainage grid   | 74        |
| 5.3      | Existing FEH catchment descriptors at high resolution                   | 79        |
| 5.4      | Proposed new catchment descriptors                                      | 87        |
| 5.5      | Automatic estimation of design flows for all points along rivers        | 89        |
| 5.6      | Closing comments on high-resolution catchment descriptors               | 94        |
| <b>6</b> | <b>Recommendations for further work</b>                                 | <b>95</b> |
|          | <b>References</b>   | <b>97</b> |

|   |   |
|---|---|
| <b>List of abbreviations</b>  | <b>105</b>  |
| <b>Appendix A: Palaeoflood techniques and data in UK river catchments</b>                               | <b>107</b>  |
| <b>Appendix B: Statistical methods for the inclusion of historical data in flood frequency analysis</b> | <b>133</b>  |
| <b>Appendix C: Feasibility study for the development of a Local and Historical Flood Data Archive</b>   | <b>179</b>  |
| Table 2.1   | Factorial standard error of flood peak estimates at ungauged sites 16   |
| Table 4.1   | Summary of local data types considered in this project 49   |
| Table 4.2   | Comparison of regression equations based on channel width 59  |
| Table 4.3   | Comparison of QMED prediction from catchment descriptors and from channel width 61  |
| Table 4.4   | Comparison of QMED prediction from catchment descriptors and from channel width 63  |
| Table 4.5   | Comparison of performance of 3 regression equations for QMED 69   |
| Table 4.6   | Recommendations for a new system to deliver local data to practitioners 73  |
| Table 5.1   | Catchment descriptors used by the FEH and ReFH methods 79   |
| Table 5.2   | Comparison of FEH and high-resolution catchment descriptors at selected gauging stations 82   |
| Table A.1   | Riverine sedimentary environments in the UK where event-scale palaeoflood records have been reconstructed 114   |
| Table B.1   | Estimated L-CV from different procedures 176  |
| Table C.1   | Recommendations for the implementation of the new system 181  |
| Table C.2   | Data groups 188   |
| Table C.3   | User roles 198  |
| Table C.4   | Risks and mitigation 208  |
| Table C.5   | Option recommendations 209  |
| Figure 2.1  | High-level condition tree for assessing uncertainty in flood hydrology 3  |
| Figure 2.2  | Condition tree for handling interacting sources of uncertainty 4  |
| Figure 2.3  | Error in $\ln(QMED_{CD})$ without donor transfer 8  |
| Figure 2.4  | Error in $\ln(QMED_{CD})$ with transfer from 1 donor 9  |
| Figure 2.5  | Error in $\ln(QMED_{CD})$ with transfer from 6 donors 10  |
| Figure 2.6  | Factorial standard error in gauged estimates of QMED 14   |
| Figure 2.7  | Factorial standard error of FEH statistical estimate versus return period for 637 essentially rural catchments that are suitable for QMED estimation 17 |
| Figure 2.8  | Sampling error in at-site $Q_T$ estimate for less permeable ( $BFIHOST < 0.65$ ) and more permeable ( $BFIHOST \geq 0.65$ ) catchments 18               |
| Figure 2.9  | Schematic representation of the ReFH2 model 20  |
| Figure 2.10   | Relationship between $QMED_{AMAX}$ and $QMED_{ReFH2}$ 21  |
| Figure 2.11   | Spatial distribution of $\ln$ -error (a) and relative error (b) in $QMED_{ReFH2}$ 22  |
| Figure 2.12   | Scatterplots of $\ln$ -error against catchment descriptors 23   |
| Figure 2.13   | Spatial distribution of error <sub>Cini</sub> 24  |
| Figure 3.1  | Improvement in the estimate of QMED as a result of substituting gauged BFI for BFIHOST 30   |
| Figure 3.2  | Use of circular diagrams to summarise and distinguish flood regimes 32  |
| Figure 3.3  | Link between Letcombe Bassett daily mean flow and Gramps Hill groundwater level 36  |
| Figure 3.4  | Winter 2013 to 2014 flows in the Ems at Westbourne, West Sussex 47  |
| Figure 4.1  | Historical data example (River Wear at Durham) 53   |
| Figure 4.2  | Comparison of bankfull channel width from the RHS and GW datasets 58  |
| Figure 4.3  | Revised regression of QMED on channel width 59  |
| Figure 4.4  | Spatial distribution of gauging stations in the BCW dataset, showing which method gave the best prediction of QMED 62                                   |
| Figure 4.5  | Scatter plot of FEH catchment descriptors and BCW 64  |
| Figure 4.6  | Relationship between BCW and catchment area 65  |
| Figure 4.7  | Comparison of predicted and observed QMED from 2 regression models 67   |
| Figure 4.8  | Comparison of predicted and observed QMED from 2 regression models 68   |
| Figure 4.9  | Change in QMED between the FEH and the BCW + CDs regressions 70   |
| Figure 5.1  | Location of hydrometric area 75 in north Cumbria 76   |
| Figure 5.2  | DTM with portions of some catchment boundaries shown as straight lines in areas of uniform flow direction 78  |
| Figure 5.3  | Flow direction grid with some catchment boundaries shown as straight lines in areas of uniform flow direction 78  |
| Figure 5.4  | Reason for catchment boundary differences for station 75001 83  |
| Figure 5.5  | Reason for catchment boundary differences for station 75006 84  |
| Figure 5.6  | Reason for catchment boundary differences for station 75007 85  |
| Figure 5.7  | Reason for catchment boundary differences for station 75017 86  |
| Figure 5.8  | Illustration of donor station used for adjusting QMED at each studied location in HA75 91   |
| Figure 5.9  | Illustration of adjustment factor for QMED at each studied location in HA75 92  |
| Figure 5.10   | Illustration of growth factor for the 100-year flood at each studied location in HA75 93  |
| Figure A.1  | Flood sediment units showing (a) in-channel bedforms, (b) channel margin accretion and (c) floodplain depositional sites 112                            |
| Figure A.2  | Flood recording environments 114  |

|             |  |     |
|-------------|--|-----|
| Figure A.3  | Flood units at Low Prudhoe, River Tyne   | 116 |
| Figure A.4  | Cut-off fill flood units from Bloody Inches, River Tay   | 118 |
| Figure A.5  | Decadal frequency of lichen-dated boulder berms in upland areas of England and Wales   | 120 |
| Figure A.6  | Relative flood magnitude based on average boulder berm b-axis measurements   | 121 |
| Figure A.7  | 3,700-year record of major floods (events plotted equate to $\leq 3\%$ of present annual exceedance probability) in the upper Severn Wales: (A) comparison with mean age of Irish bog oaks (Leuschner et al. 2002), a regional hydroclimate proxy; (B) comparison with reconstructed winter NAO index (Trouet et al. 2009); and (C) comparison with the pollen record from Carneddau in the headwaters of the Rhiw catchment (Walker 1993) | 122 |
| Figure B.1  | Historical data example (River Wear at Durham), showing a total of $k = 6$ historical events (red bars) above the perception threshold $X_0$ (dashed red line), recorded across the $h = 154$ year-long historical period  | 134 |
| Figure B.2  | RMSE for the shape parameter as a function of the systematic record length, for the different historical record length using the likelihood estimation method  | 141 |
| Figure B.3  | RMSE for the shape parameter as a function of the systematic record length, for the different historical record length using the (P)PWM estimation method  | 141 |
| Figure B.4  | Bias for the shape parameter as a function of the systematic record length, for selected ratio of historical to systematic record lengths  | 142 |
| Figure B.5  | Standard error for the shape parameter as a function of the systematic record length, for selected ratio of historical to systematic record lengths  | 143 |
| Figure B.6  | RMSE for the scale parameter as a function of the systematic record length, for the different historical record length using the likelihood estimation method  | 144 |
| Figure B.7  | RMSE for the scale parameter as a function of the systematic record length, for the different historical record length using the (P)PWM estimation method  | 144 |
| Figure B.8  | RMSE for Q5 as a function of the systematic record length for selected ratio of historical to systematic record lengths  | 145 |
| Figure B.9  | RMSE for Q200 as a function of the systematic record length for selected ratio of historical to systematic record lengths  | 146 |
| Figure B.10 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 10 using the likelihood estimation method   | 147 |
| Figure B.11 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 30 using the likelihood estimation method   | 148 |
| Figure B.12 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 46 using the likelihood estimation method   | 148 |
| Figure B.13 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 76 using the likelihood estimation method   | 149 |
| Figure B.14 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 10 using the (P)PWM estimation method   | 150 |
| Figure B.15 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of using the (P)PWM estimation method  | 151 |
| Figure B.16 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 46 using the (P)PWM estimation method   | 151 |
| Figure B.17 | Standard error for the $\log(QT)$ as a function of the return period for a systematic sample size of 76 using the (P)PWM estimation  | 152 |
| Figure B.18 | Estimated shape parameter when using systematic data only or systematic and historical data. Left: estimates obtained with (P)PWM method. Right: estimates obtained with Likelihood Method   | 153 |
| Figure B.19 | Relationship between the shape parameter and the flood frequency curve (GLO distribution)  | 153 |
| Figure B.20 | Proportion of cases in which the estimate of the shape parameter using systematic and historical data is larger than when using only systematic data in simulated records for which no event in the systematic record exceeded the perception threshold $X_0$ using the likelihood estimation method   | 154 |
| Figure B.21 | Proportion of cases in which the estimate of the shape parameter using systematic and historical data is larger than when using only systematic data in simulated records for which at least one event in the systematic record exceeded the perception threshold $X_0$ using the likelihood estimation method   | 155 |
| Figure B.22 | Proportion of cases in which the estimate of the shape parameter using systematic and historical data is larger than when using only systematic data in simulated records for which at least one event in the systematic record exceeded the perception threshold $X_0$ using the (P)PWM estimation method   | 156 |
| Figure B.23 | RMSE for the shape parameter when using known values of the historical data or when only the number of perception threshold exceedances is known as a function of the systematic sample size   | 157 |
| Figure B.24 | RMSE for the $\log(QT)$ with a systematic sample size of 10 when using the known values of the historical data or when only the number of perception threshold exceedances is known  | 158 |
| Figure B.25 | RMSE for the $\log(QT)$ with a systematic sample size of 76 when using the known values of the historical data or when only the number of perception threshold exceedances is known  | 159 |
| Figure B.26 | RMSE for the shape parameter as a function of the number of historical events missing in the historical record using the likelihood estimation method  | 160 |
| Figure B.27 | RMSE for the shape parameter as a function of the historical events missing in the historical record using the likelihood estimation method  | 161 |
| Figure B.28 | RMSE for the $\log(QT)$ for a systematic sample size of 46 and an historical period covered of (A) $h = 230$ and (B) $h = 2,300$   | 162 |
| Figure B.29 | RMSE for the shape parameter as a function of the historical events missing in the historical record using the (P)PWM estimation method  | 163 |
| Figure B.30 | RMSE for the shape parameter as a function of the historical events missing in the historical record using the (P)PWM estimation method  | 164 |
| Figure B.31 | RMSE for the shape parameter as a function of the systematic record length for selected historical record lengths  | 166 |
| Figure B.32 | RMSE for the scale parameter as a function of the systematic record length for selected historical record lengths  | 167 |

|             |   |     |
|-------------|---|-----|
| Figure B.33 | RMSE for the shape parameter as a function of the systematic record length, for selected historical record lengths                          | 168 |
| Figure B.34 | Bias for the shape parameter as a function of the systematic record length for selected historical record lengths                           | 169 |
| Figure B.35 | RMSE for the $\log(Q_T)$ for a systematic sample size of 46 and an historical period covered of $h = 46$ (left panel) and 920 (right panel) | 170 |
| Figure B.36 | RMSE for the shape parameter as a function of the systematic record length for selected historical record lengths                           | 171 |
| Figure B.37 | Bias for the shape parameter as a function of the systematic record length, for selected historical record length 171                       | 171 |
| Figure B.38 | Annual maxima peak flow values used in the case study   | 173 |
| Figure B.39 | L-moment diagram for systematic data only showing at-site and pooled estimates  | 174 |
| Figure B.40 | L-moment diagram when including historical data: at-site and pooled estimates are shown.  | 175 |
| Figure B.41 | Estimated flood frequency curves according to different methods   | 176 |
| Figure C.1  | Hierarchy of data usefulness  | 191 |
| Figure C.2  | Conceptual scheme of the system and interaction with users  | 193 |
| Figure C.3  | Illustrative example of specialisation of data types  | 194 |

# 1 Introduction

Flood frequency estimates are an essential part of flood risk management. They are an important ingredient of many important decisions, informing the cost-effectiveness, design and operation of flood defences, flood mapping and planning decisions in flood risk areas. They also inform the National Flood Risk Assessment, the setting of insurance premiums and long-term investment planning.

Methods described in the Flood Estimation Handbook (FEH) published in 1999, and many subsequent updates, are considered the industry standard for flood estimation in the UK. They are used extensively by hydrologists from both the public and private sectors.

Flood frequency estimates – also known as design flood estimates – are associated with many sources of uncertainty. These hydrological uncertainties often constitute the most uncertain component in any flood study. Uncertainty can lead to difficulty in having confidence in the outputs of studies, whether these are for investment planning, insurance, asset design, development planning or other purposes. As a result, there is considerable benefit to be gained from any reduction in the uncertainty of flood frequency estimation.

There are many supplementary sources of information that can help to refine estimates of design floods and potentially reduce uncertainty. Examples include long-term flood history, river level records, photographs of floods and information obtained from field visits.

These and similar types of information are defined as ‘local data’. The FEH Local research project aimed to:

- quantify the uncertainty of design floods estimated from FEH methods
- develop procedures and guidance for incorporating local and historical data into flood estimation to reduce uncertainties

The primary objective of this report is to describe the reviews and research carried out during the FEH Local project.

Another output from the project was a document giving guidance to practitioners on how to estimate uncertainty in flood frequency and how to find and incorporate local data. The practitioner guidance, ‘Using Local Data to Reduce Uncertainty in Flood Frequency Estimation’, will be disseminated early in 2017.

This report aims to avoid duplication with the practitioner guidance and so is intended mainly for those with an interest in the background to the methods presented in the guidance.

## 2 Uncertainty in design flood estimation

### 2.1 Protocols for managing uncertainty

There is a substantial body of literature relating to the assessment and management of uncertainty in flood risk analysis, including flood hydrology. Among many papers and books on the subject, relevant and comprehensive reviews have been published by Beven (2010), covering a broad spectrum of environmental and hydrological modelling, and Beven and Hall (2014), with a focus on flood risk management.

There has been often intense debate within the scientific literature about the appropriate treatment of uncertainty in flood hydrology and the wider fields of environmental risk management and natural hazards. In many papers, the debates touch on philosophical arguments about the meaning of information, knowledge and data. Other papers focus on deeply technical aspects of statistical models for estimation errors, or explore numerous alternative formalisms for representing various different types of uncertainties.

Although some of the academic debates may appear esoteric to the hydrologist working in practice, it is through this intellectual effort that a coherent approach to uncertainty management is gradually emerging. Most studies set out a classification or typology of uncertainty that acknowledges a distinction between uncertainties that should be treated as inherently unavoidable ‘randomness’ and others that are regarded as ‘knowledge errors’ that could in principle be corrected. However, the way in which specific sources of uncertainty are classified can vary between studies and is a matter of debate. The statistical models used for quantification also vary widely.

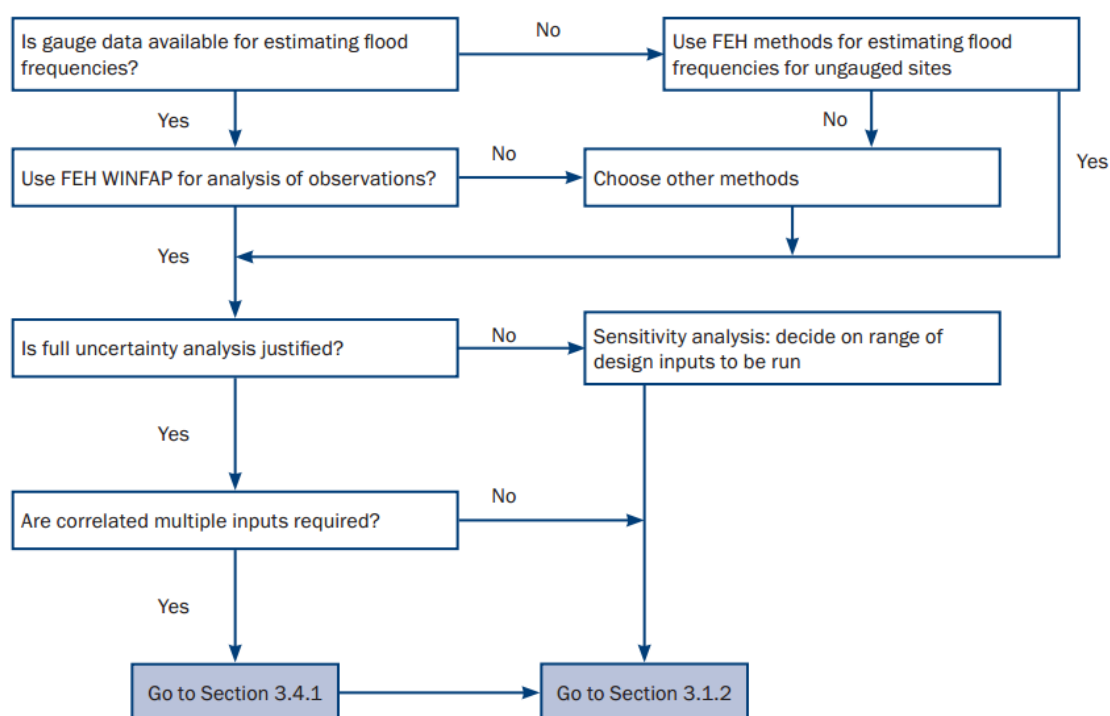
One of the most important insights to be gained from the research is that no single, unique ‘uncertainty method’ has emerged for flood hydrology. In illustrating this point, Kjeldsen et al. (2014b) demonstrate several alternative methods available for the analysis of uncertainty in flood estimation, each one scientifically credible and based on published theory, and each leading to somewhat different results. The results, of course, depend on assumptions made in the analysis, and it is not always, if ever, obvious which method should be preferred. Indeed, the precise assumptions made in a specific analytical method may not be apparent or even acceptable to an individual hydrologist. Kjeldsen et al. (2014b) therefore advocated that estimates of uncertainty should be accompanied by a description of the analysis, specifying which parts of the modelling system are considered as known without uncertainty, and which parts are assumed to contribute to the uncertainty.

As a consequence of ambiguities about the choice of method for quantifying uncertainty, it is inevitable that some subjectivity will be inherent in any realistic assessment. This mirrors some of the issues about choice of hydrological model (for example, statistical or rainfall-runoff), where objective evidence is usually mixed with a handful of judgement calls. This subjectivity was taken as a starting point in a recent attempt to formalise good practice guidance for managing uncertainty and published as a CIRIA report (Beven et al. 2014). The good practice guide was designed to cover all aspects of flood mapping and therefore includes flood estimation as a central element. The motivation for these good practice guidelines was to encourage clarity and transparency in expressing and agreeing the judgements involved in any management of uncertainty. The CIRIA good practice guidelines are presented as a sequence of



steps to record assumptions and choices made in assessing the uncertainty for future reference.

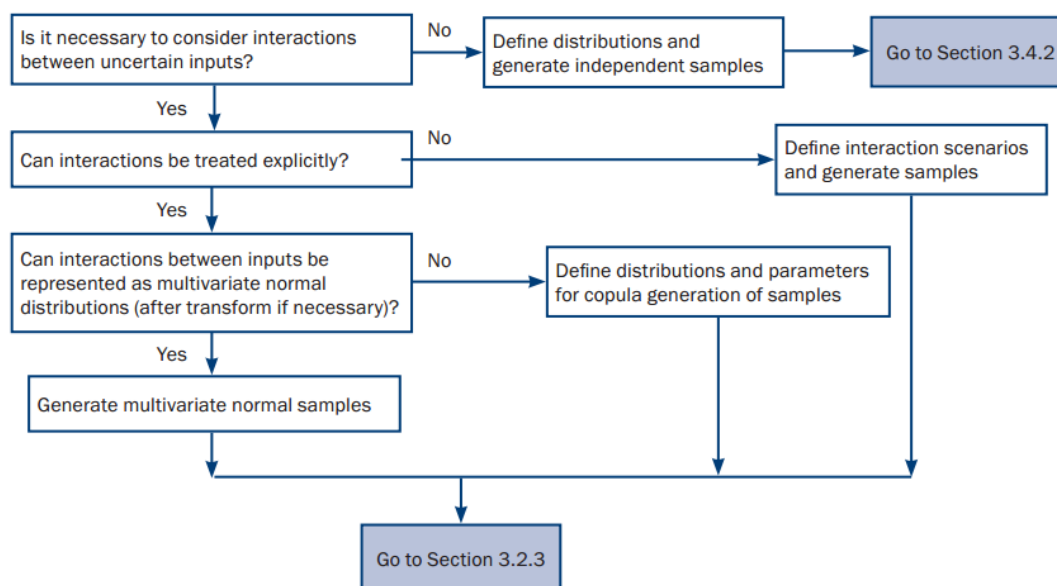
It is important that the approach to management of uncertainty is proportional to the expected value of any particular decision or scheme being considered. The CIRIA guidelines suggest different levels of analysis within a single framework of condition (or decision) trees, within which the assumptions made at each stage are recorded for later evaluation. At the highest level, the condition tree for uncertainty estimation in flood frequency analysis (Figure 2.1) acts as a filter to guide analysis in the most straightforward situations towards relatively simple, accessible methods while offering an alternative path towards potentially more complex and costly approaches where these may be necessary.



**Figure 2.1 High-level condition tree for assessing uncertainty in flood hydrology**

Notes: Figure 3.1 in Beven et al. (2014), where Section 3.4.1 refers to guidelines for the analysis of interacting sources of uncertainty and Section 3.1.2 refers to the most generic level of guidance.

A more complex analysis may be appropriate if there is a need for a comprehensive uncertainty assessment (related to a high value investment, for example) and where there are interactions between multiple sources of uncertainty, which may demand a more complicated analysis to be robust against scientific challenge. In this case, the CIRIA good practice guidelines would lead to a lower-level condition tree as shown in Figure 2.2.



**Figure 2.2 Condition tree for handling interacting sources of uncertainty**

Notes: Figure 3.9 in Beven et al. (2014)

The CIRIA good practice guidelines are arguably the first attempt to formalise a management protocol for uncertainty in flood estimation. They include case studies, showing how the condition trees lead to an audit trail of analytical decisions. In some respects, the guidelines reflect a particular view of uncertainty developed since the 1990s through the work of the lead author Keith Beven and numerous collaborators. Naturally, there may be different views expressed by other authors, although it should be noted that the guidelines were developed through the Flood Risk Management Research Consortium (I and II) projects, which successfully integrated a broad academic consortium with representatives from leading industry partners. Therefore, the guidelines can be considered a collegiate effort that reflects views from within industry.

## 2.2 Applications of uncertainty assessment in the flood management industry

The CIRIA good practice guidelines are not prescriptive and could be regarded in some respects as technically ambitious for the industry in general. This ambition is appropriate if flood risk management is to be based on sound evidence to support confident decision-making, but the development of good practice remains an ongoing process.

The most established responses to uncertainty in ‘knowledge’ tend to be scenario analysis. A well-known example of this is the use of several agreed scenarios of greenhouse gas emissions to develop projects of the future climate.

Consideration of the uncertainty associated with inherent randomness tends to lead to probabilistic methods, as set out in the following sections. An example already applied in English flood risk management is the RASP (Risk Assessment for Strategic Planning) approach for modelling the reliability of flood defences.

The assessment of uncertainty can feed into decision-making in many ways, including approaches based on formal decision theory (for a review in the context of UK climate

change planning, see Ranger et al. 2010) or probabilistic risk-based analysis, which is attempted to some extent in the Environment Agency's National Flood Risk Assessment (NaFRA) methodology and also within catastrophe models used in the insurance sector.

The questions asked of hydrological analysis may need to evolve to bring uncertainty into the decision-making process. For example, it would not be very helpful to conclude that the designed crest level for a flood defence should be  $Y \pm x$  mm, but it might be helpful to ask a question along the lines of 'What is the cheapest design we are x% confident has a y% likelihood of not being overtopped?'.

## 2.3 Estimating uncertainty

In the UK, flood frequency estimation is most commonly conducted according to the 2 methods published in the FEH (Institute of Hydrology 1999) and its subsequent major updates (Kjeldsen 2007, Environment Agency 2008).

A recent development has been the release of an update to the Revitalised Flood Hydrograph (ReFH) method, ReFH2. One novel aspect of ReFH2 is the improved modelling of urban catchments, as described by Kjeldsen et al. (2013). Urban and rural areas within the catchment of interest are modelled separately. Other aspects of the ReFH2 method have not yet been fully published. The ReFH2 software was released by Wallingford HydroSolutions in 2015.

It is common in many disciplines to express uncertainty in terms of variance ( $s^2$ ) or its square root, standard deviation ( $s$ ). The sample variance describes the scatter of the sample values ( $x_1, \dots, x_n$ ) around the sample mean  $m$  and is estimated as:

$$s^2 = ((x_1 - m)^2 + \dots + (x_n - m)^2) / (n - 1) \quad (\text{equation 2.1})$$

For a normally distributed variable,  $X$ , with mean  $\mu$  and variance  $\sigma^2$ , the intervals  $\mu \pm \sigma$  and  $\mu \pm 2\sigma$  contain approximately 68% and 95% of the distribution probability, respectively. Hence, approximately 68% and 95% of all observations of a sample  $\mathbf{x}$  from  $X$  should fall within the  $m \pm s$  and  $m \pm 2s$  intervals, respectively, where  $m$  and  $s^2$  are the estimated sample mean and variance. Samples that exhibit behaviour different from this general rule might be representative of non-normal distributions, or be affected by outliers that undermine the estimation of the sample mean and sample variance.

Many common variables in hydrology are assumed to be normally distributed after taking a log transformation. For log-transformed variables in general, the 68% confidence interval for the mean of the distribution is  $x \pm s$ , where  $x$  is the sample average and  $s$  is the standard deviation of the sample average. The confidence interval for the original variable can be taken as approximately  $e^x e^{\pm s}$ . The 68% confidence interval for  $x$  may then also be expressed as  $[x/fse; x \cdot fse]$  and the 95% confidence interval as  $[x/fse^2; x \cdot fse^2]$ , where  $fse$  denotes the factorial standard error.

QMED, the index flood corresponding to the location parameter of the Generalised Logistic (GLO) distribution, is generally assumed to be log-normally distributed.

The uncertainties connected to flood frequency estimation within the FEH methods are discussed in the next section. The discussion concerns only sampling uncertainty, that is, the uncertainty that arises from using records of finite length to estimate the design flood events. Another potentially important source of uncertainty is related to the choice of flood frequency model, which can have important implications for design floods at

higher return periods. Although the GLO distribution is typically assumed as the default choice, Kjeldsen and Prosdocimi (2015) have shown that the choice between a GLO model and a Generalised Extreme Value (GEV) model is, in many cases, not always obvious. In addition, the effects of measurement errors and uncertainties in rating curves are not discussed, although these are likely to have an effect on the performance of any model.

With respect to FEH Local, one of the important controls on uncertainty is the role of local information in refining generalised estimates.

## 2.4 Uncertainty in the FEH statistical method

### 2.4.1 Overview

In the improved FEH statistical method, the uncertainty of a design flood estimate  $\hat{Q}$  results from uncertainty around the estimation of the index flood  $\hat{\mu}$ , estimation of the growth curve  $z_T$  and the covariance between the two (Equation 2.2; from Kjeldsen 2015). Thus,

$$var(\hat{Q}) = z_T^2 var(\hat{\mu}) + \mu^2 var(\hat{z}_T) + 2 \mu z_T cov(\hat{\mu}, \hat{z}_T) . \quad (\text{equation 2.2})$$

With some assumptions, uncertainty in QMED ( $var(\hat{\mu})$ ) can be expressed simply through  $fse$ . However, uncertainties in the growth curve and covariance are more complicated to determine and need to consider the effects of dependence between annual maxima series in a pooling group (Kjeldsen and Jones 2004, 2006).

### 2.4.2 Uncertainty in QMED

Methods for estimating the uncertainty (expressed as  $fse$ ) for estimates of QMED are reviewed in this section for a number of cases that typically occur in practical use of the FEH methods.

#### *At ungauged sites*

If no annual maximum data are available at the site of interest, an initial estimate of QMED for a rural or assumed rural catchment is made from an equation combining 4 catchment descriptors (Equation 2.3).

$$QMED_{CD} = 8.3062 AREA^{0.851} 0.1536^{(1000/SAAR)} FARL^{3.4451} 0.0460^{BFIHOST^2} \quad (\text{equation 2.3})$$

where AREA is the catchment area (km<sup>2</sup>), SAAR is the standard annual average rainfall (mm), FARL is the FEH index of flood attenuation due to reservoirs and lakes, BFIHOST is the Base Flow Index (BFI) estimated from soil type and the subscript CD indicates an estimate obtained from catchment descriptors only.

The uncertainty in this estimate of QMED is a function of the model structure and of the variability of the catchment descriptors, and hence is fixed. The structure of the

equation considers the model residuals normally distributed when modelling  $\ln(\text{QMED})$ . It is therefore appropriate to discuss uncertainty in terms of  $fse$ , which is reported as 1.431 for Equation 2 (Environment Agency 2008). As  $1.431^2 \approx 2$ , there is a 5% probability that the true median of annual maxima at a site is either less than half or more than double the value of QMED found from Equation 2.3.

In recognition of the high degree of uncertainty associated with estimates of QMED obtained using regression relationships such as Equation 2.3, it is considered best practice to use data transfer from gauged local donor catchments whenever possible (NERC 1975, Institute of Hydrology 1999, Environment Agency 2008). However, simplified versions of the FEH donor transfer procedure, in which either the nearest gauge or that with the most similar catchment is selected as a donor, would lead to an increase in the uncertainty of the adjusted QMED estimate (Kjeldsen and Jones 2007). Kjeldsen and Jones presented a revised version where the ratio between the observed and predicted QMED at the donor site is adjusted according to the geometric distance between the centroids of the donor and subject sites.

Donor transfer is a mechanism that exploits spatial correlations in the residual of modelled QMED: the ratio of observed to modelled QMED is found at the nearest gauged catchment and this, raised to a power term  $\alpha$ , is used to improve the modelled estimate of QMED at the catchment of interest. After donor transfer, the  $fse$  of the adjusted QMED at the ungauged catchment is given by:

$$fse = \exp[(s^2[1 - \alpha^2] + s_d^2\alpha^2)^{0.5}] \quad (\text{equation 2.4})$$

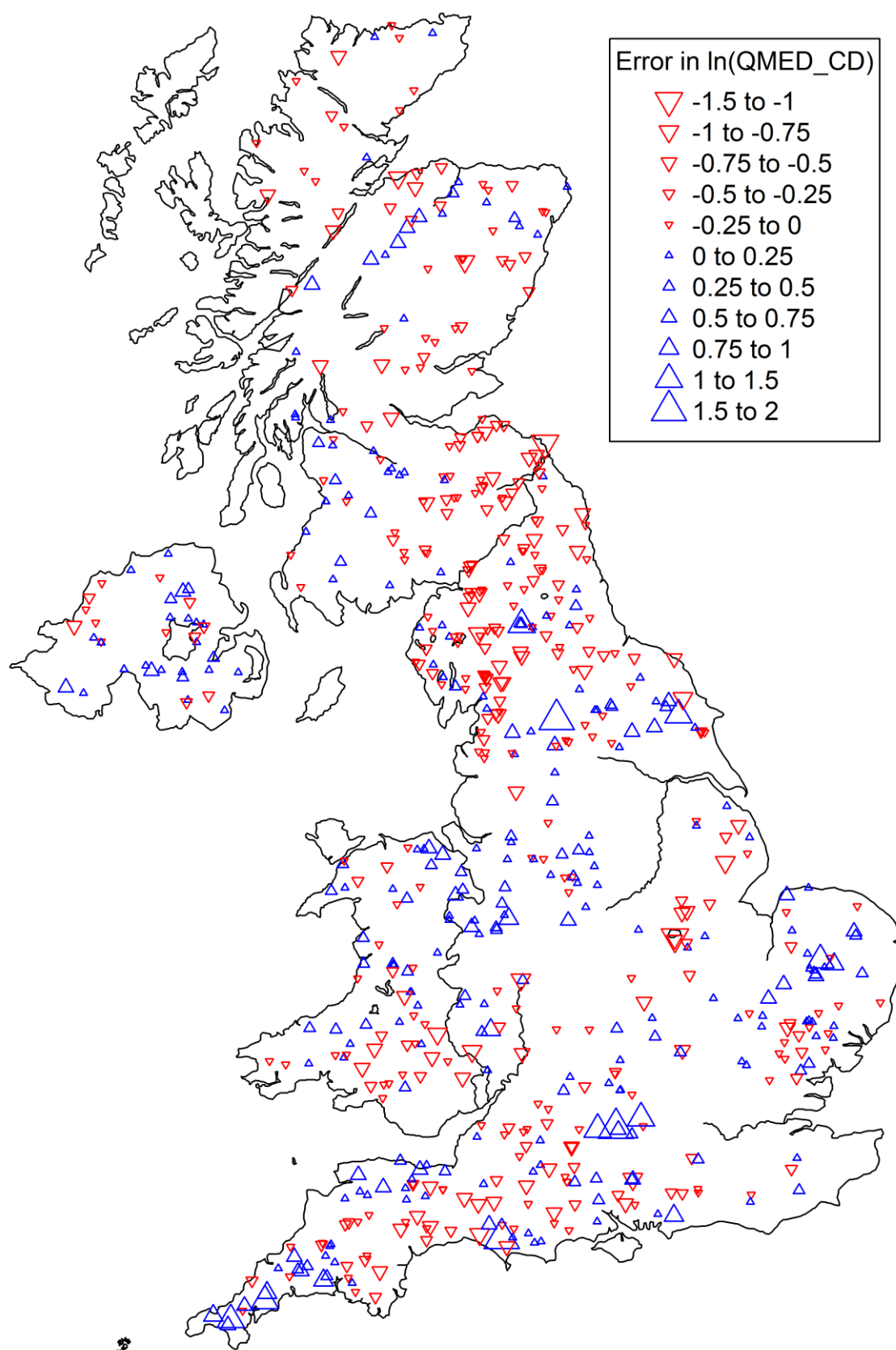
where  $s$  is the standard error of the QMED regression model (Equation 2.3) residuals and  $s_d$  is the standard error of sampled  $\ln(\text{QMED})$  at the gauged donor site. The term  $\alpha$  is related to distance between the centroids of the donor and the catchment of interest, and is approximately 0.5 at 1.5km and 0.1 at 18.7km. The formula used to compute  $\alpha$  is given in Environment Agency (2008).

In cases where data from more than one donor site are used to correct the initial catchment descriptor-based estimate of QMED, the  $fse$  value for the adjusted QMED value is:

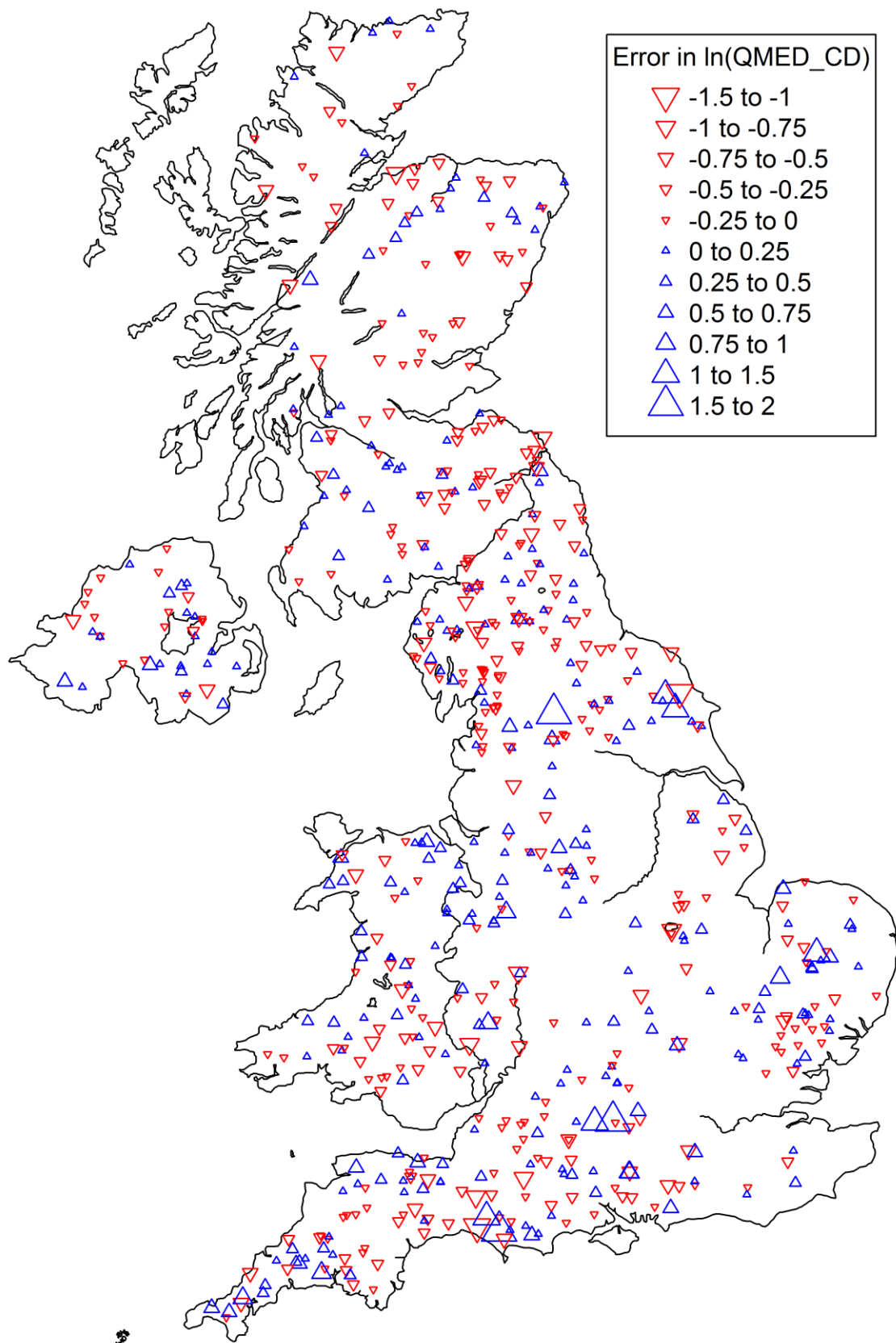
$$fse = \exp[(s^2 - \mathbf{b}^T \mathbf{\Omega}^{-1} \mathbf{b})^{0.5}] \quad (\text{equation 2.5})$$

where  $\mathbf{b}$  is a vector containing the covariance between the subject site and each donor site, and  $\mathbf{\Omega}$  is a matrix containing the covariance between the model error for each pair of donor catchments. Further details regarding the use of multiple donor sites are given in Kjeldsen et al. (2014a).

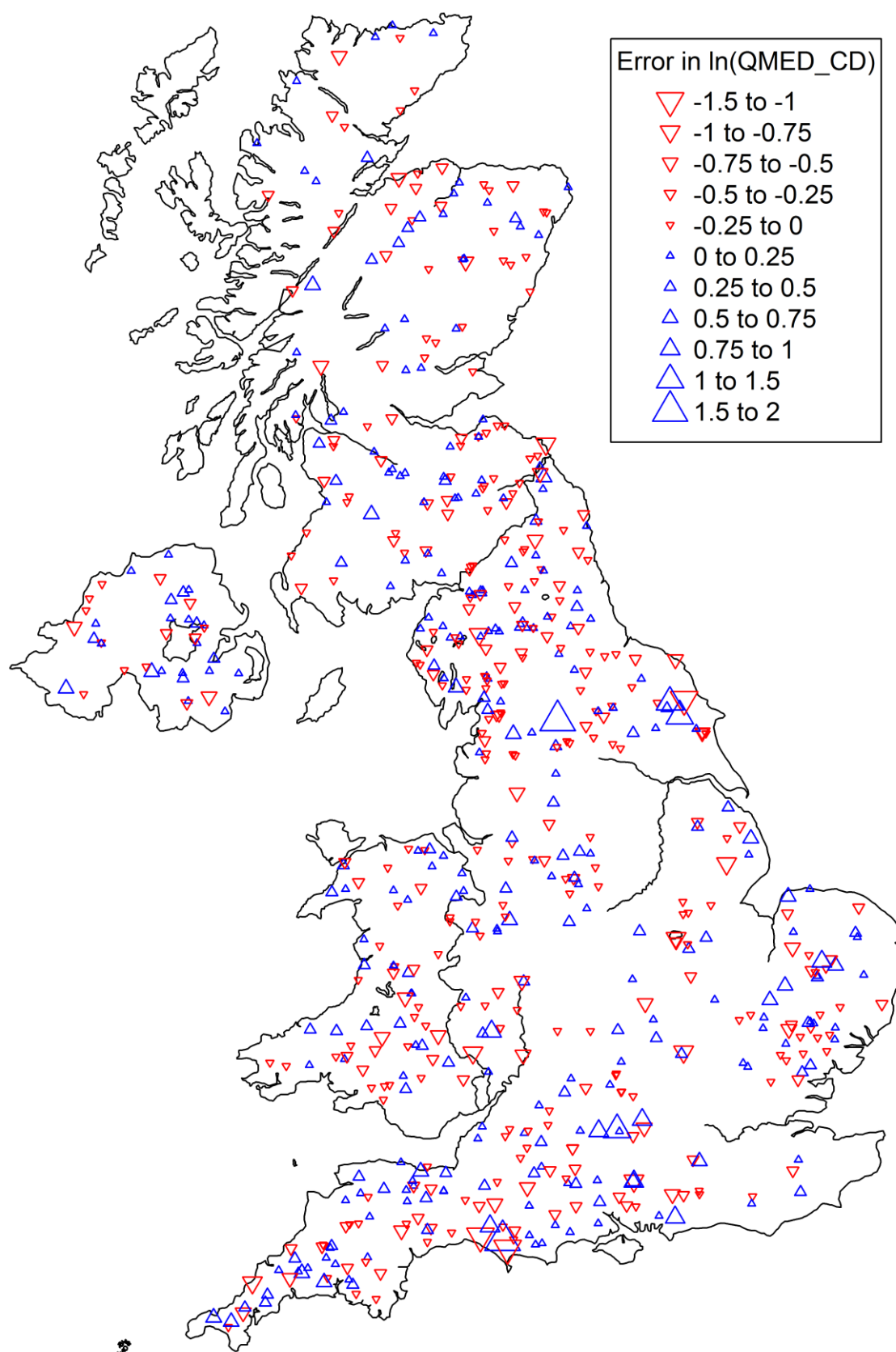
Figures 2.3, 2.4 and 2.5 plot the error in estimated  $\ln(\text{QMED}_{\text{CD}})$  with 0, 1 and 6 donors, respectively (selected entirely on the basis of proximity to the subject catchment). These figures show some spatial clustering of positive and negative errors, which are counteracted by introducing information from nearby sites.



**Figure 2.3** Error in  $\ln(\text{QMED}_{\text{CD}})$  without donor transfer



**Figure 2.4** Error in  $\ln(\text{QMED}_{\text{CD}})$  with transfer from 1 donor



**Figure 2.5** Error in  $\ln(\text{QMED}_{\text{CD}})$  with transfer from 6 donors



## At gauged sites

If annual maxima are available at the catchment of interest, the best estimate of QMED is generally given directly by the median value in the annual maximum record. The *fse* value of QMED estimated in this way can also be found directly from the gauged data, under the common assumption that annual maxima series follow a GLO distribution. The *fse* value of the median derived from a gauged series of annual maxima is:

$$fse = \exp[2\beta/n^{0.5}] \quad (\text{equation 2.6})$$

where  $\beta$  is the scale parameter of the GLO distribution and  $n$  is the number of annual maxima at the site (Kjeldsen and Jones 2006). The *fse* value of QMED, as estimated from gauged annual maximum records, is less than 1.10 at over 80% of the 963 sites in the National River Flow Archive (NRFA) peak flow dataset<sup>1</sup>.

The constant value of 2 in Equation 2.6 is derived from a first-order approximation whose accuracy depends on the degree of non-linearity in the quantile function (that is, GLO distribution). The value 2 is always approached asymptotically as series length increases, but can vary greatly for shorter records. Kjeldsen (2015) shows that this constant value is far less applicable when short data series with large negative skewness are involved. Coupled with the fact that the term  $\beta$  in Equation 2.6 depends upon L-CV and L-skewness, it is theoretically possible for Equation 2.6 to give a value of *fse* exceeding that of the regression model for ungauged sites (which is 1.431). Some investigation of the NRFA peak flow dataset indicates that this occurs only at a single site (no. 33049, Stanford Water at Buckenham Tofts), which has a short record with high L-CV; 2 of the 7 gauged annual maxima are more than 4 times the median value, while the next 3 are within 10% of the median value.

A Monte Carlo simulation study was performed to generalise *fse* in at-site estimates of QMED. For this, 100,000 random records with lengths of 3–10, 15, 25, 50 and 100 years were sampled from GLO distributions with each combination of the following:

- Scale parameter: 0.025 to 0.8 in steps of 0.025
- Shape parameter: -0.8 to +0.8 in steps of 0.05

For each combination, the 100,000 QMED estimates were ln-transformed and *fse* was found for the ln(QMED) estimates. Record lengths of 1 and 2 annual maximum flows were not considered as it is not possible to generate a shape parameter for a series with less than 3 values.

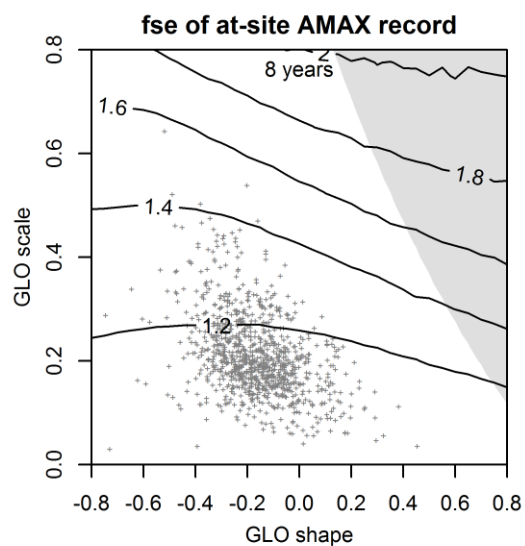
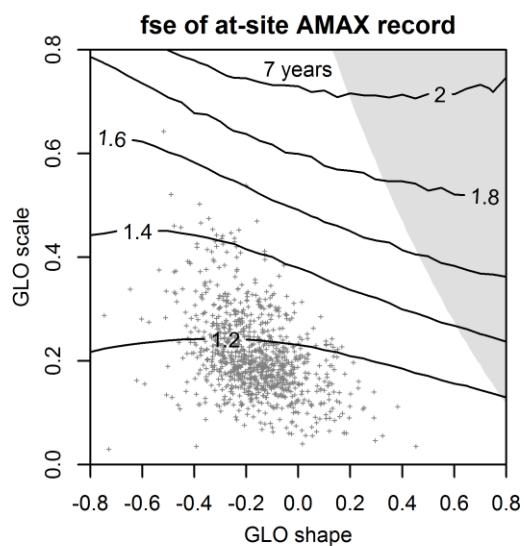
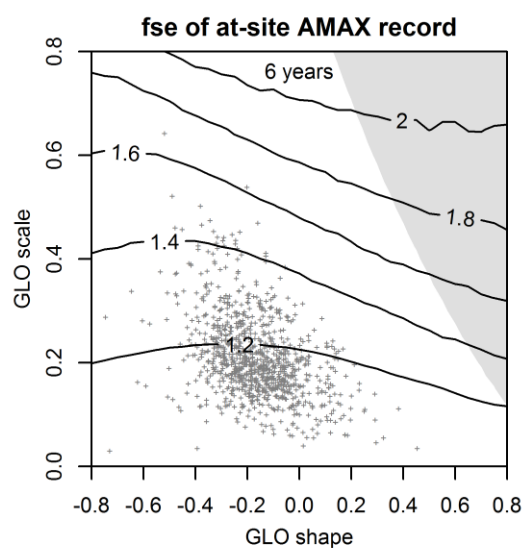
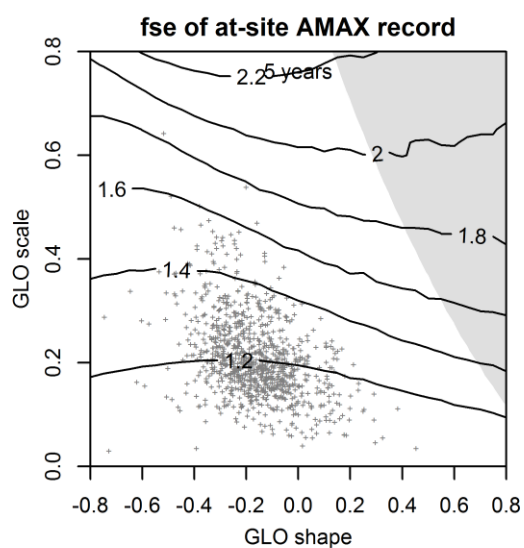
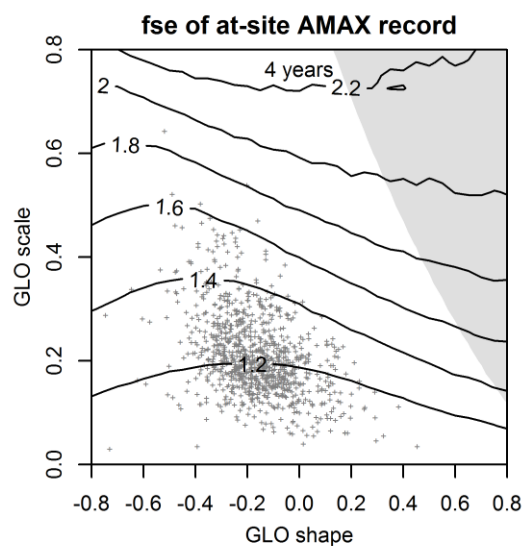
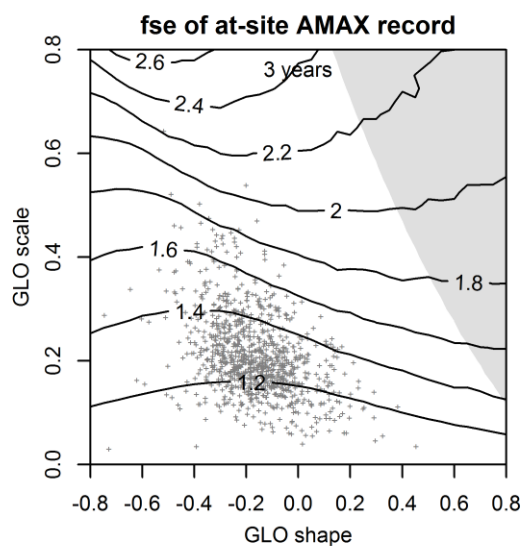
Figure 2.6 plots the results of this study, showing that even 3 annual maxima are sufficient for the accuracy of a gauged value of QMED to exceed that of one estimated from catchment descriptors at most stations. By the time the number of annual maxima reaches 6, the gauged estimate of QMED is more accurate than that from catchment descriptors at almost 100% of stations. As discussed below, in most practical situations it is preferable to use peaks-over-threshold (POT) data to estimate QMED when records are very short.

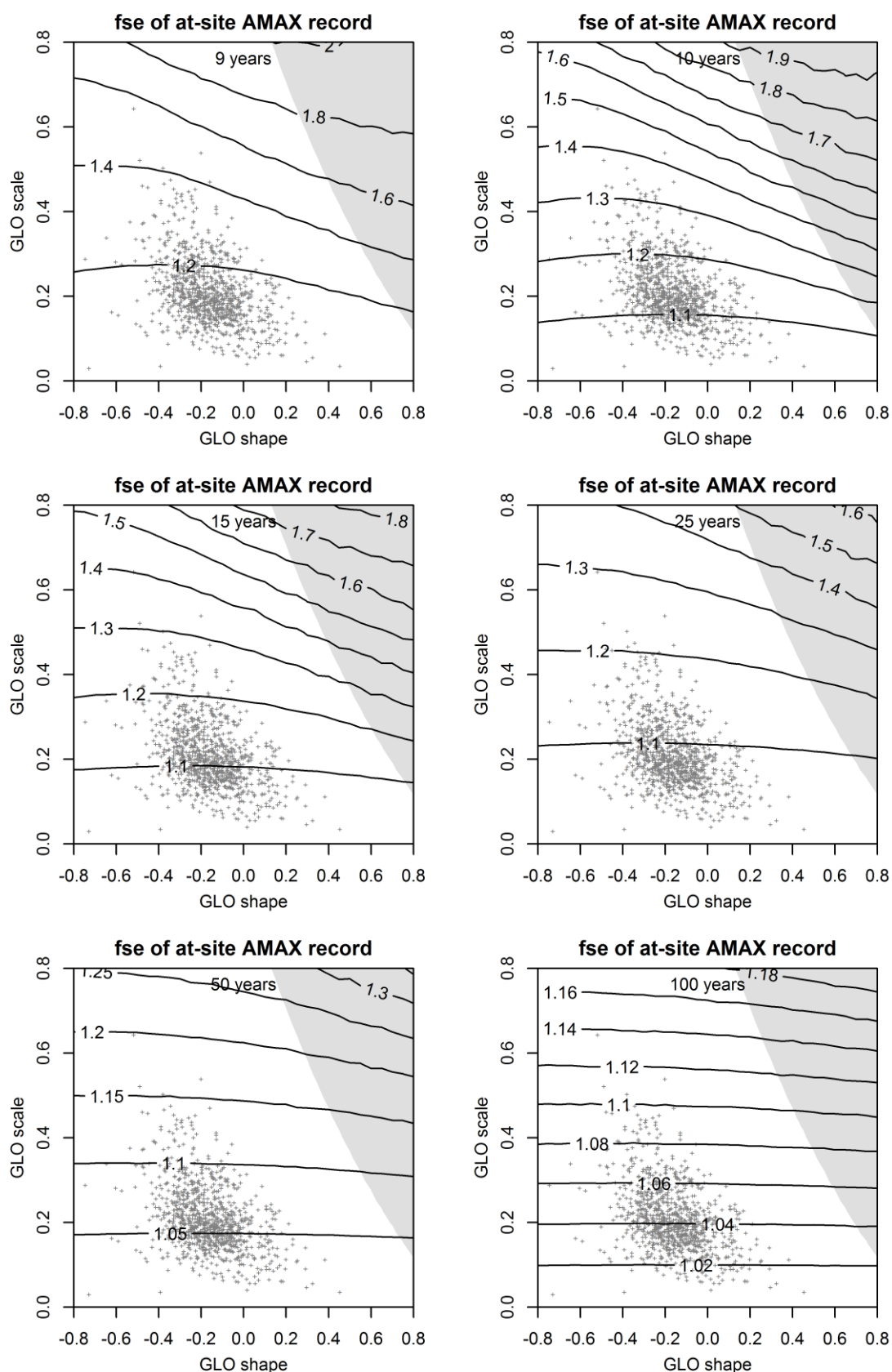
Additionally, Figure 2.6 shows that gauged records with low GLO scale parameter values generate QMED estimates with lower *fse*, tending to one as the parameter value tends to zero. If a short record has a high GLO scale value, each additional annual maximum value that can be added causes *fse* to fall relatively rapidly. Gauged

---

<sup>1</sup> <http://nrfa.ceh.ac.uk/>

records with positive GLO shape parameter values generate QMED estimates with higher fse, although the relative importance of the shape parameter value decreases for longer records. In practical terms, a record with a low GLO scale parameter corresponds to one in which all annual maxima are similar and, therefore, are all near QMED; conversely, a record with a positive GLO shape parameter features two or more tied or near-tied largest flows that are far from the median value.





**Figure 2.6 Factorial standard error in gauged estimates of QMED**

Notes: Points represent combinations of GLO scale and shape parameters found in the NRFA Peak Flow dataset. Shaded areas represent impossible combinations of parameter values, assuming that all flow data are non-negative.

### *At sites with short gauge records*

If a valid POT record exists at the site of interest, it has the potential to provide significantly more information than the equivalent annual maximum record and, therefore, a more accurate and less uncertain estimate of QMED. The method published in the FEH is to rank all flood peaks by magnitude and evaluate QMED as a weighted average of 2 consecutive values, the positions of which depend on the number of years in the POT record. Typical confidence intervals for QMED values estimated from POT data are tabulated in Volume 3 of the FEH (Table 2.2). No method is given to calculate the specific confidence intervals around a QMED value estimated from a specific POT record. Owing to the difficulty in obtaining good quality POT data, this method was not revisited during the last major update of the FEH statistical method (Environment Agency 2008) and has not been revisited otherwise.

If it is not feasible to extract POT data (for example, on a river where flow is dominated by baseflow even in flood conditions), and only a short annual maximum record is available, then it may not be appropriate to estimate QMED purely from that record owing to sampling error. One approach to finding a better estimate of QMED may be to take a weighted average of the value given by estimation from annual maximum data and the value given by the regression equation:

$$QMED = w QMED_{cfs} + (1 - w) QMED_{obs} \quad (\text{equation 2.7})$$

where  $w$  is a weighting factor given as:

$$w = s_d^2 / (s^2 + s_d^2) \quad (\text{equation 2.8})$$

The  $fse$  value of this estimate of QMED is:

$$fse = \exp[(s_d s) / (s_d^2 + s^2)^{0.5}] \quad (\text{equation 2.9})$$

Kjeldsen (2015) uses an example (site no. 56013, River Yscir at Pont-yr-Yscir) to show that just 5 annual maxima recorded at a site can reduce  $fse$  (and hence the confidence interval) around QMED by considerably more than donor transfer. This finding highlights the importance of using local data whenever available.

If the site of interest has only a very short record from a temporary station, it may be possible to extend that record by regression. Flow data from the temporary station are compared with flow data from the closest station on the same river network over the same time period and a regression relationship is derived to explain the flow at the temporary station as a function of the flow at the permanent station. This relationship is then applied to the entire flow series at the permanent station to estimate the flows at the site of the temporary station over the same time period. From this, a POT or annual maximum series can be derived and assessed using an appropriate method for a gauged site.

Record extension is only suitable if the regression has high explanatory power; the FEH suggests that it should be able to explain more than 90% of the variance in flood peaks at the subject site. In any case, the use of a regression relationship adds further

uncertainty to the sampling and measurement errors expected at the permanently gauged site.

### 2.4.3 Uncertainty in design flows for longer return periods

As a flood frequency estimate for any event other than the 2-year peak considers the use of the index flood and growth curve together, covariance between these 2 components also forms part of the uncertainty in the estimate.

#### *At ungauged sites*

Kjeldsen (2015) attempts to evaluate the uncertainties at ungauged sites associated with return periods longer than 2 years by measuring *fse* values between estimates made purely from at-site data (single-site analysis) and estimates made from pooling groups that exclude the catchment of interest (ungauged analysis). The results are shown in Table 2.1, where *fse* is shown to increase with return period and at a similar rate regardless of whether donor transfer is employed.

**Table 2.1 Factorial standard error of flood peak estimates at ungauged sites**

| Return period | <i>fse</i> (regression only) | <i>fse</i> (regression + donor) |
|---------------|------------------------------|---------------------------------|
| 2 years       | 1.47                         | 1.42                            |
| 5 years       | 1.48                         | 1.43                            |
| 30 years      | 1.52                         | 1.47                            |
| 100 years     | 1.54                         | 1.50                            |

Source: Kjeldsen (2015, Table 6)

Sampling errors in the at-site records are assumed to be independent and identically distributed with zero mean, which allows them to be accounted for using a Monte Carlo simulation described in the appendix of Kjeldsen (2015). The Monte Carlo simulation by itself may therefore also be used to estimate the sampling error and confidence intervals associated with a T-year event in analysis of a single site.

The full expression used for calculating *fse* is:

$$fse = \exp \left[ \sqrt{\frac{1}{m} \sum_{i=1}^m (\ln \hat{Q}_T - \ln Q_T)^2 - \frac{1}{m} \sum_{i=1}^m \text{var}\{\varepsilon^2\}} \right] \quad (\text{equation 2.10})$$

where *m* is the number of catchments,  $\hat{Q}_T$  is the FEH statistical estimate,  $Q_T$  is the single-site estimate and  $\varepsilon$  is the sampling error in the at-site record.

Additional work conducted as part of the FEH Local project has applied this same method of calculating *fse* to 637 members of the NRFA peak flow dataset that are both

suitable for QMED estimation and essentially rural ( $URBEXT_{2000} < 0.03$ ),<sup>2</sup> excluding 2 sites for which the digital catchment area is known to be a poor estimate of drainage area.

This additional work considered a wider range of return periods (2 to 2,000 years) and donor transfer from 0, 1, 2 and 6 donors, and accounted for sampling error by a Monte Carlo simulation with 100,000 rather than 10,000 repeats. The results of this work are plotted in Figure 2.7 and suggest that  $fse$  can be estimated for any return period from 2 to 2,000 years via quadratic equations:

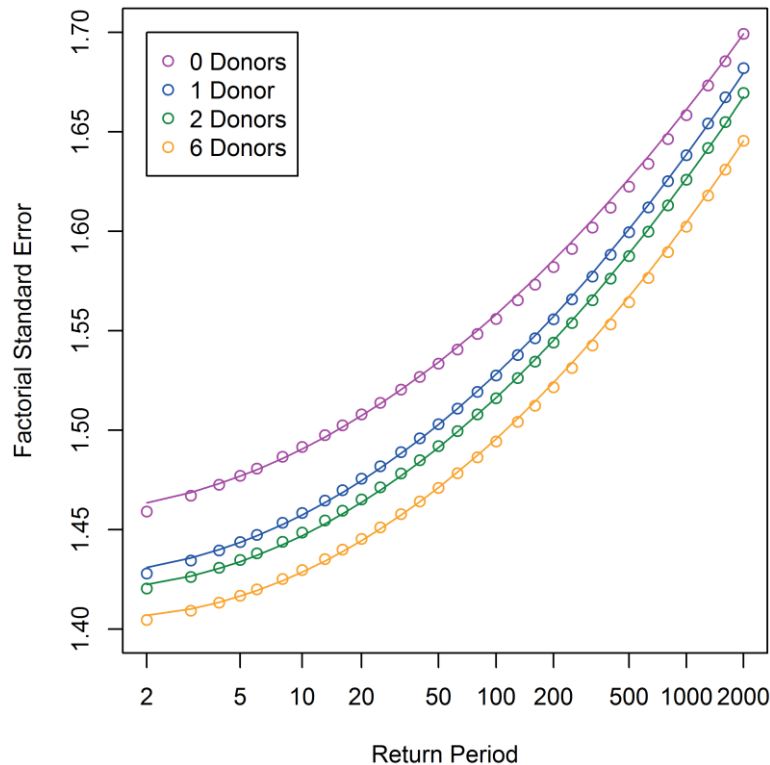
$$fse_0 = 1.461 + 0.0055y + 0.0034y^2 \quad (\text{equation 2.11})$$

$$fse_1 = 1.429 + 0.0041y + 0.0038y^2 \quad (\text{equation 2.12})$$

$$fse_2 = 1.421 + 0.0028y + 0.0039y^2 \quad (\text{equation 2.13})$$

$$fse_6 = 1.406 + 0.0011y + 0.0040y^2 \quad (\text{equation 2.14})$$

where  $y$  is the Gumbel reduced variate,  $-\ln(-\ln(1-1/T))$  and the subscripts after  $fse$  correspond to the number of donors used to adjust the index flood estimate. The values of  $fse$  shown in Figure 2.7 are broadly consistent with those found by Kjeldsen (2015), but with a smaller improvement in performance attributable to donor transfer and possibly faster growth in  $fse$  with increasing return period. Both these factors can probably be attributed to differences between this dataset and Kjeldsen's. Figure 2.7 clearly shows that the incremental advantage of using 6 donors rather than one is similar to or better than the incremental advantage of using one donor rather than zero, with the greatest advantage at longer return periods.



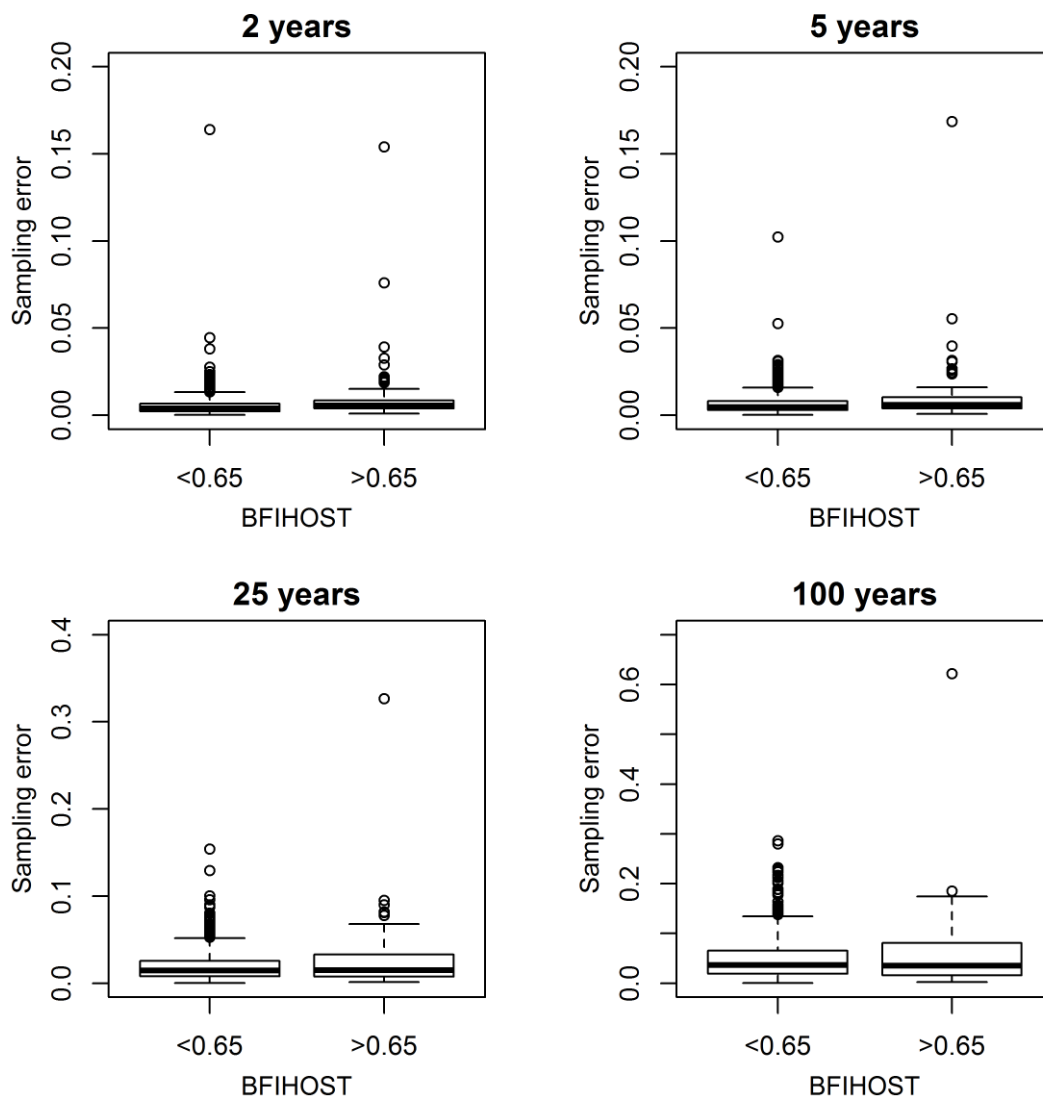
**Figure 2.7 Factorial standard error of FEH statistical estimate versus return period for 637 essentially rural catchments that are suitable for QMED estimation**

<sup>2</sup> URBEXT is the FEH index of fractional urban extent.

Figure 2.8 shows a boxplot of sampling error for  $T = 2, 5, 25$  and 100 years for the full set of 637 catchments, divided into a lower BFIHOST group with 550 members and a higher BFIHOST group with 87 members. The boxplot shows the interquartile range (IQR) of sampling errors, while the whiskers extend to 1.5 times the IQR. The thick line inside the box marks the median.

Figure 2.8 shows that catchment permeability has almost no effect on values of sampling error up to and including the median, while having only minimal effect on all but the highest outliers. This is corroborated by the position of the top whisker in the lower BFI group versus the higher BFI group, and the resulting number and position of outliers in each case.

Although there is little difference between sampling errors in low and high BFIHOST catchments, it is possible that  $fse$  for a  $Q_T$  estimate in an ungauged high BFIHOST catchment may be larger than in an ungauged low BFIHOST catchment. This is because the FEH statistical methodology was developed using a calibration dataset very similar to the current NRFA peak flow dataset, in which less permeable catchments greatly outnumber more permeable catchments. In general, the methods might give slightly less satisfactory results for catchments whose properties are underrepresented in the calibration dataset.



**Figure 2.8 Sampling error in at-site  $Q_T$  estimate for less permeable ( $BFIHOST < 0.65$ ) and more permeable ( $BFIHOST \geq 0.65$ ) catchments**



This study also considered the relationship between  $fse$  and return period for subsets of the NRFA peak flow dataset in defined URBEXT<sub>2000</sub> groups: 0.03–0.06, 0.06–0.15, 0.15–0.30 and 0.30–0.60. However, the difference in behaviour between different groups was too great, and the number of catchments per group too small (79, 83, 29 and 19, respectively), to rule out significant sampling uncertainty.

### *At gauged sites (enhanced single-site method)*

The enhanced single-site method (Environment Agency 2008) combines an at-site annual maximum record with the FEH statistical pooling procedure. QMED is estimated as the median of at-site annual maxima, while the pooled L-moment ratios  $t_2$  and  $t_3$  contain contributions from the at-site L-moments and those of the hydrologically similar sites necessary to bring the combined record length to at least 500 years. The at-site data are considered more important than data at other catchments, so are weighted differently from the L-moment ratios in the rest of the pooling group.

It was initially proposed to measure uncertainty in the enhanced single-site method, using the same procedure as for the FEH statistical method. However, this was not done: the enhanced single-site and at-site estimation methods are highly dependent, as the largest component of the enhanced single-site estimate is the at-site data. Consequently, the difference between the at-site and enhanced single-site estimates ( $\ln \hat{Q}_T - \ln Q_T$ ) is regularly less than the sampling error in the at-site data  $\epsilon$ . It is evident that sampling error must not be double counted, but it is not clear how to avoid doing so. Further work on accounting for sampling error, or a different procedure, is required for uncertainty in the enhanced single-site method to be evaluated.

## 2.5 Uncertainty in the ReFH2 method

### 2.5.1 Approach to assessing uncertainty in ReFH2

The FEH provides an alternative method of flood estimation based on the use of the ReFH rainfall-runoff model in combination with a set of design storm inputs and initial conditions as described by Kjeldsen (2007). The design method has been updated to include a revised structure and new parameter estimation equations (Wallingford HydroSolutions 2016) and is now known as ReFH2.

Owing to the complexity of the modelling procedures that underlie ReFH2, estimating the uncertainty in the resultant design hydrographs is not straightforward. Uncertainty in ReFH2 estimates may derive from a number of sources including:

- measurement error in the flood and rainfall event data used to calibrate the model
- sampling error
- model error related to the form of the rainfall-runoff model

To this can be added the uncertainty of the design rainfall inputs (which are again subject to measurement, sampling and model errors), the design initial conditions and the assumptions on which the design method is based (for example, the choice of duration and the use of unrealistic rainfall profiles).

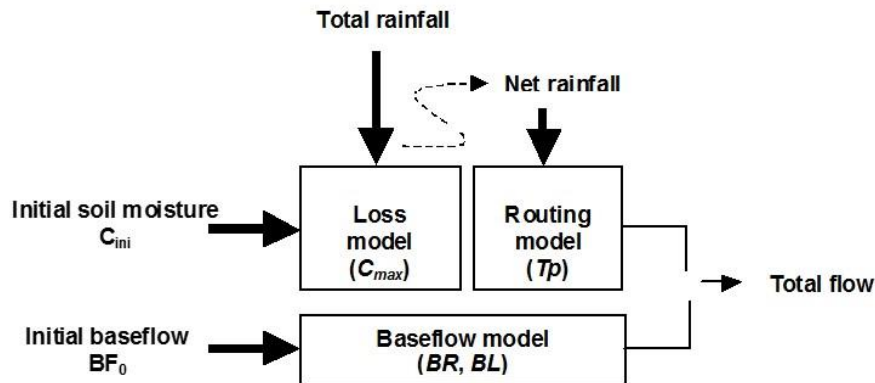
It is therefore difficult to estimate the uncertainty of the design flood estimates produced by ReFH2. However, a number of practical approaches could be worthwhile and these are discussed in the following sections.

## 2.5.2 Uncertainty in ReFH2 parameters

The ReFH2 design method requires 4 parameters:

- maximum soil moisture capacity ( $C_{max}$ )
- unit hydrograph time to peak ( $T_p$ )
- baseflow lag (BL)
- baseflow recharge (BR)

Two initial conditions are also required – initial soil moisture content ( $C_{ini}$ ) and initial baseflow ( $BF_0$ ) – as illustrated schematically in Figure 2.9.



**Figure 2.9 Schematic representation of the ReFH2 model**

Source: Wallingford HydroSolutions (2016, Figure 1)

The values of these parameters and initial conditions can be estimated from gauged flow records if available, but more often, they will be estimated via parameter estimation equations based on FEH catchment descriptors. Further details can be found in the ReFH2 technical guidance (Wallingford HydroSolutions 2016). The design storm duration ( $D$ ) is also required and this is estimated from an equation taken from the Flood Studies Report (FSR) rainfall-runoff method (NERC 1975).

The regression equation for each of the model parameters has an associated factorial standard error ( $fse$ ). This implies that the log-transformed regression equations have a mean error of zero and a standard error of  $\ln(fse)$ . Hence, uncertainty in each input parameter value could be simulated by adding random samples from a normal distribution with mean zero and standard error equal to  $\ln(fse)$  to the calculated regression value. It would be necessary to run ReFH2 a large number of times, in each instance adding a random sample from a distribution describing the regression error to each of the 4 calculated parameter values. Uncertainty in the initial conditions could be simulated by using the 'randomised' values of  $T_p$  and  $C_{max}$  to recalculate  $D$ ,  $C_{ini}$  and  $BF_0$ .

This type of approach could be used to provide an indication of how variable the ReFH2 estimates are once uncertainty in the parameters has been taken into account. The main issue is that there is a strong correlation between the parameters and also a strong correlation between the parameter correlations and catchment descriptors (and possibly also between parameter uncertainty and catchment descriptors). This makes it very difficult to identify all the sources of uncertainty and to assess the influence of the catchment descriptors on estimation error.

A similar approach has been implemented in the ReFH1 module of the Flood Modeller Pro software. The 'Probabilistic ReFH' tool is intended to indicate the uncertainty in modelled flows and water levels as a result of uncertainty in the ReFH1 model parameters, initial conditions and design storm duration. The software produces a set of 33 equally probable combinations of the model parameters and design inputs. A set of 33 hydrographs can then be produced from these combinations, which can be ranked in order of peak flow, total volume or time to peak. Since all these hydrographs are intended to be equally probable, ranking them in order is used to derive a percentage exceedance.

### 2.5.3 Comparisons of modelled QMED with peak flow data

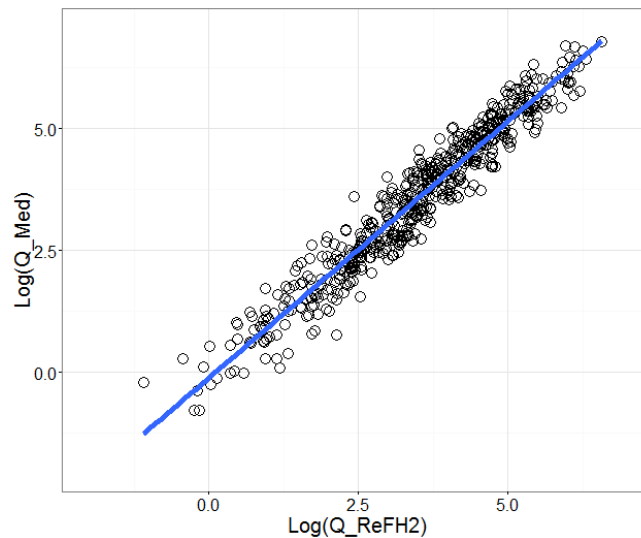
Following discussions with the project team, it was agreed that a more practical analysis might prove to be more worthwhile than a detailed exploration of the possible uncertainty in the ReFH2 parameters and design inputs. A comparison between QMED estimated from ReFH2 ( $QMED_{ReFH2}$ ) and the median of AMAX flow peaks from the NRFA peak flow dataset ( $QMED_{AMAX}$ ) was therefore carried out. The analysis used data and model predictions for 571 catchments selected to match the following criteria:

- essentially rural catchments ( $URBEXT2000 < 0.03$ )
- flow regime not influenced by reservoirs or lakes ( $FARL > 0.9$ )
- classified as suitable for QMED estimation

Linear regression was used to investigate the relationship between the logarithms of measured and modelled QMED for each catchment (Figure 2.10). Assuming that the prediction variance is dominated by the variance of the regression model residuals ( $s^2 = 0.153$ ), it is possible to estimate the factorial standard error ( $fse$ ) as presented in Kjeldsen (2015) as:

$$fse_{ReFH2} = e^s = 1.48 \quad (\text{equation 2.15})$$

The  $fse_{ReFH2}$  value of 1.48 is slightly higher than that from the regression model used in the FEH procedure to estimate the QMED from catchment descriptors ( $fse_{reg} = 1.431$ ). However, it is important to recognise that the ReFH2 method is not directly calibrated to QMED values.



**Figure 2.10 Relationship between  $QMED_{AMAX}$  and  $QMED_{ReFH2}$**

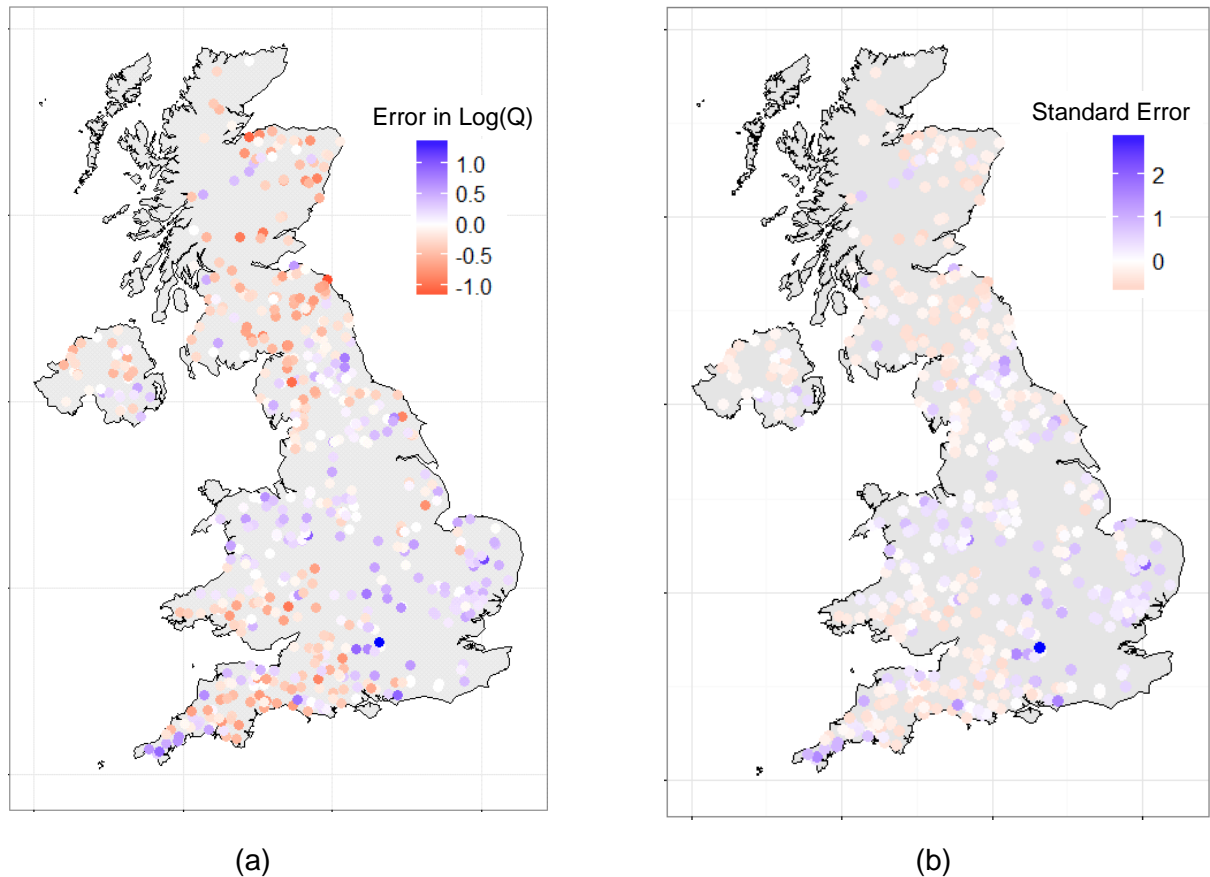
Figure 2.11 presents the spatial distribution of the ln error (Equation 2.16) and the relative error (Equation 2.17) in  $QMED_{ReFH2}$  computed for each catchment.

$$\ln \text{ error} = \ln(QMED_{ReFH2}) - \ln(QMED_{AMAX}) \quad (\text{equation 2.16})$$

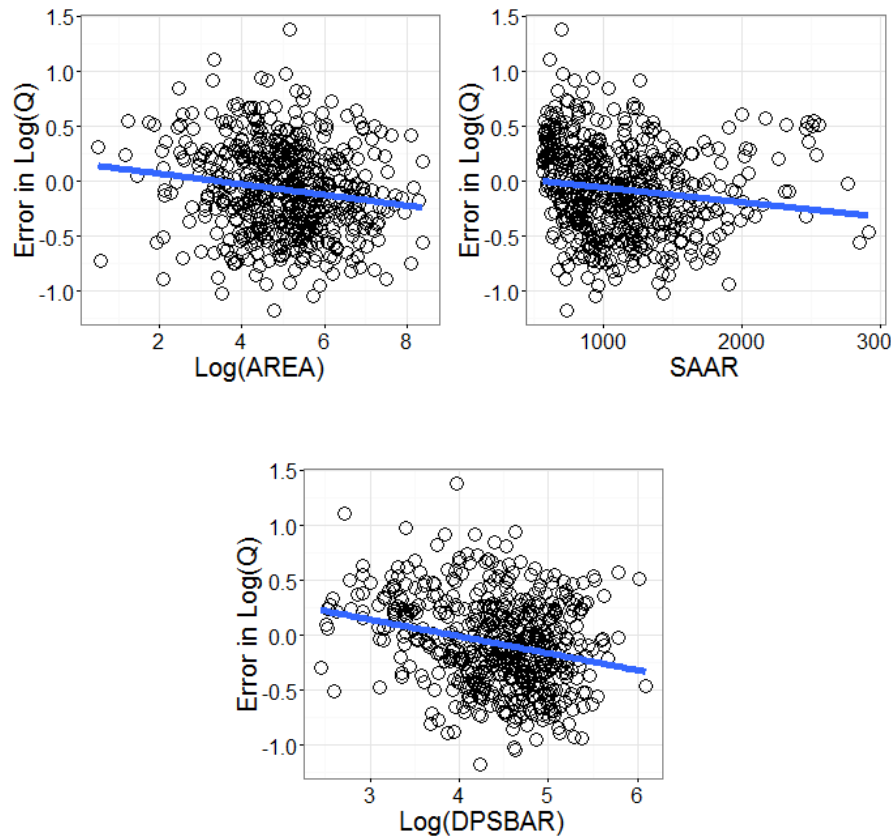
$$\text{relative error} = (QMED_{ReFH2} - QMED)/QMED \quad (\text{equation 2.17})$$

An outlier catchment can be seen in Figure 2.11 in central southern England. This is the Pang at Pangbourne (39027), a permeable catchment affected by groundwater abstraction. The maps show that the ReFH2 model tends to slightly underpredict measured QMED in the north and west parts of the UK and to slightly overpredict in the south-east.

Further analysis was undertaken to investigate possible relationships between ln-error and catchment descriptors, and example scatter plots are presented in Figure 2.12. As AREA, SAAR and DPSBAR (FEH index of mean drainage path slope) increase, there is a tendency for the ReFH2 design model to underpredict QMED.



**Figure 2.11** Spatial distribution of ln-error (a) and relative error (b) in  $QMED_{ReFH2}$



**Figure 2.12 Scatterplots of In-error against catchment descriptors**

#### 2.5.4 Estimation of $C_{ini}$ in ReFH2

The estimation of an appropriate value of  $C_{ini}$ , the initial soil moisture content, is a critical step in the ReFH2 design package. A new model for  $C_{ini}$  is incorporated in ReFH2 based on the estimation of initial soil moisture for the 2-year return period event (the 5-year event was adopted in the original ReFH design package). The estimation procedure was as follows (Wallingford HydroSolutions 2016).

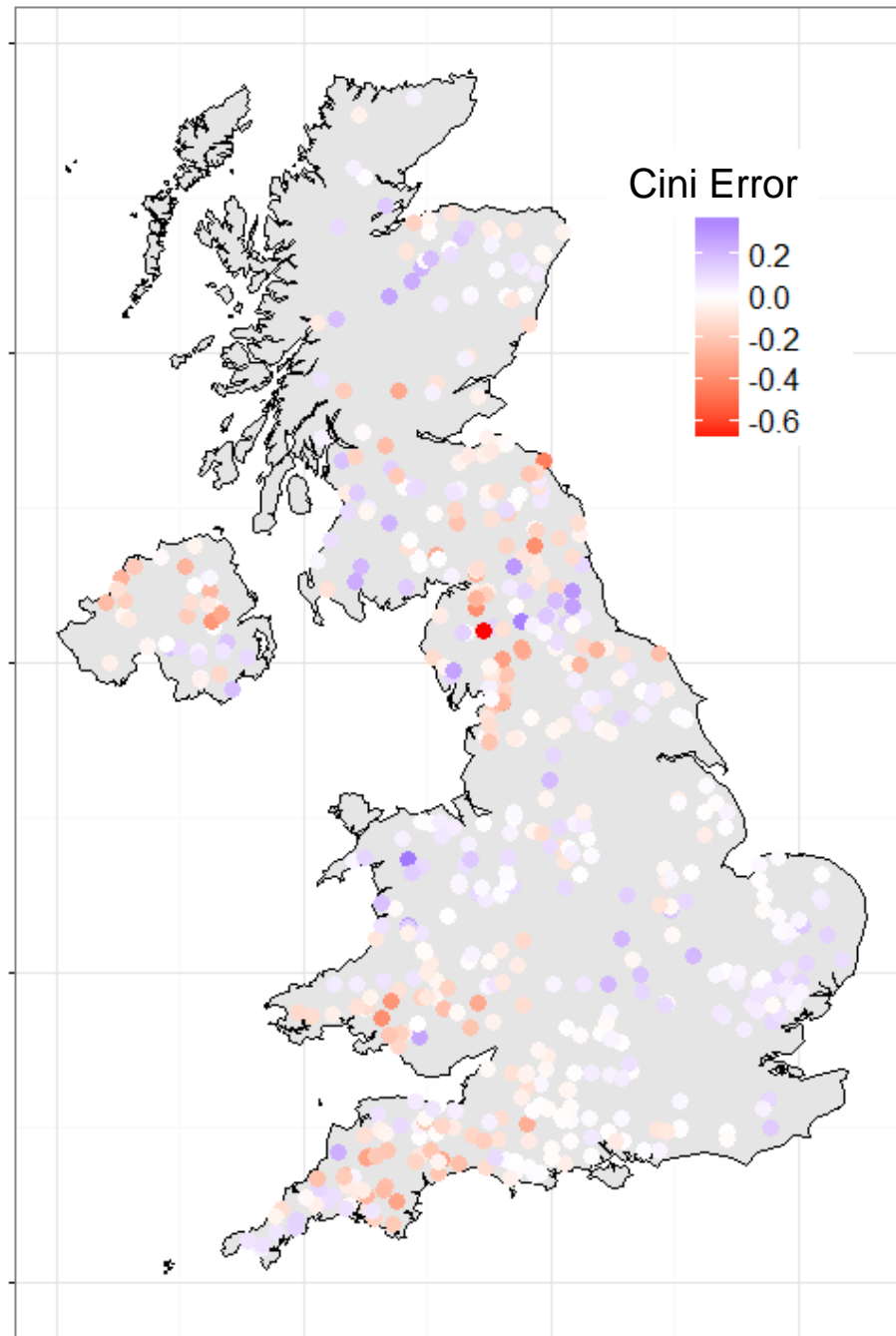
- The 2-year design storm was estimated using the FEH99 rainfall depth–duration–frequency model in conjunction with the recommended duration.
- ReFH was run with design package estimates including  $BF_0$ .
- The value of  $C_{ini}$  required to calibrate the ReFH estimate of the 2-year peak flow to the value of QMED estimated directly from the gauged record ( $C_{ini\_ideal}$ ) was identified.
- The resultant set of  $C_{ini\_ideal}$  values across all catchments was used to develop a model for estimating  $C_{ini}$  from catchment descriptors.

An initial analysis of the difference between  $C_{ini\_ideal}$  and  $C_{ini}$  estimated from catchment descriptors ( $C_{ini\_model}$ ) was carried out for the set of 571 catchments described above. Figure 2.13 presents the spatial distribution of the error $_{C_{ini}}$  as defined in Equation 2.18:

$$error_{C_{ini}} = C_{ini\_model} - C_{ini\_ideal} \quad (\text{equation 2.18})$$

The model performs reasonably well across the UK. Error values vary between -0.1 and 0.1 in central and southern parts of the UK, and there is a slight tendency towards underestimation in the south-west and overestimation in the north.

There appears to be some scope for introducing a method of data transfer similar to that used in the FEH statistical method, where  $C_{ini\_ideal}$  values at gauged sites could be used to correct the value of  $C_{ini\_model}$  derived from catchment descriptors at ungauged sites.



**Figure 2.13** Spatial distribution of error<sub>Cini</sub>

# 3 Review of local data for flood frequency estimation

## 3.1 Introduction

This chapter reviews existing research and practice on the use of local data for flood frequency estimation. It does not examine the development of new or revised methods, which are covered in Chapter 4.

## 3.2 Definition of local data

The best estimates of design floods are almost always made from analysis of local records of accurately measured peak flows. In the UK, FEH methods provide opportunities for incorporating flood peak data from nearby donor sites when they are not available at the site of interest. The approaches for doing so have been refined through research over the past decade, as described, for example, by Kjeldsen and Jones (2007) and Kjeldsen et al. (2014a).

For the purposes of the present project, local data are defined as information additional to this primary data source. The project scoping document uses the following definition:

‘Local flood data can be thought of as information on flood events within a relatively small geographical area that can be used to complement traditional flood estimates (typically derived using the statistical method and river flow records from the HiFlows–UK database)’.

Local data can be useful on catchments with little or no high quality flood flow data, but in many situations, they can also be expected to significantly improve flood estimates even where flood flow data are available. For example, information on flood history can often be helpful in setting a 30- or 40-year gauged flow record in a longer context. Local data can be used to adjust generalised estimates obtained from FEH methods, with the aim of reducing uncertainty in design flows.

## 3.3 Types of local data

### 3.3.1 A broad classification

Merz and Blöschl (2008) proposed 3 types of extra information that may be usefully incorporated in flood frequency analysis in addition to any flood peak data at the site of interest. These are:

- Temporal information (that is, information on flood behaviour before or after the observed period)
- Spatial information (that is, information from other catchments)
- Causal information (that is, information on the generating mechanisms of floods)

The applicability of these 3 categories was considered with reference to the current study. The findings can be summarised as follows.

- Additional temporal information is available from nearby long records or from longer-term flood history or palaeoflood evidence.
- The FEH statistical approach already provides procedures for incorporating spatial information.
- Causal information can be incorporated to some extent when applying the ReFH method.

Although existing methods make some allowance for these types of information, Merz and Blöschl point out that there are subtleties in hydrological processes that are difficult to capture by formal methods but may be amenable to hydrological reasoning, which often needs to be site-specific. They make a plea for a shift away from solving the estimation problem to hydrological understanding. Numerous examples of information expansion, all from catchments in Austria, are given by Merz and Blöschl (2008). Several are potentially relevant to the current study, as discussed below.

Rosbjerg et al. (2013) provide a synthesis of recent worldwide research into the prediction of floods in ungauged basins. A short section outlines approaches that include proxy data on flood processes, including information on historical floods and recent post-flood information.

With regard to UK methods of design flood estimation, Kjeldsen (2015) concludes that:

‘There is still considerable uncertainty associated with predictions made in ungauged catchments, but that use of local data can help to reduce the uncertainties’.

Some uses of local data relate less to a specific source of substitute data than to recognition of (and allowance for) the specific setting of a subject site and the configuration of its catchment.

## 3.4 Perspectives on the use of local data

### 3.4.1 A long history and a range of perspectives

The importance of incorporating additional data has long been recognised, both in the research literature and in guidance for practitioners. Merz and Blöschl (2008) quote a US Geological Survey Water Supply Paper by Slade (1936) which states:

‘... the statistical method, in whatever form (graphic or analytical) is an entirely inadequate tool in the determination of flood frequencies. When used in conjunction with non-statistically inferred data, however, it may attain a high order of precision’.

The potential for flood estimates to be enhanced through the use of local data has not always been heeded. The same authors point out that most peer-reviewed publications focus on solving the statistical estimation problem, rather than giving guidance on how to incorporate hydrological reasoning.

Hydrologists from different backgrounds will bring different perspectives on the use of local data. Researchers may tend to place more value on statistical orthodoxy, developing and favouring methods that give the best performance in statistical tests carried out at a national scale. Practitioners may have a more site-specific focus and bring experience in applying more subjective techniques that involve hydrological reasoning. Both of these perspectives were included in the research for the FEH Local project, helped by the fact that the project team included representatives from both the academic and applied fringes of flood hydrology.



Local data are more than 'local knowledge', which is often attributed to lay people such as farmers and residents. However, there is an important role for local knowledge in contributing to the understanding of a river system. However, there can be a tension in flood management between expert knowledge and local knowledge, as discussed by Haughton et al. (2015) with reference to the East Yorkshire floods of 2007 and the Somerset Levels inundation in 2014.

### **3.4.2 Influence of catchment size**

Rivers draining large catchments are more likely to have gauged information directly upstream or downstream of the subject site. In these cases, the choice of donor site in QMED adjustment may be so obvious as to be self-selecting. There is also an expectation that local peculiarities in soil type, terrain or land use tend to average out over larger catchments. This general expectation does not apply in:

- highly permeable catchments
- in catchments where soils and land use are notably heterogeneous
- in catchments where major floodplains intervene

Some examples of this are given in Section 0.

The situation is different in many small and very small catchments. There is often no obvious donor catchment. Indeed, there may be no credible donor at all. Many researchers and practitioners may resort to a default method based on catchment descriptors. If the flood estimation problem is sufficiently important, experienced practitioners can be expected to prioritise understanding of the subject catchment and seek markers of its flood behaviour.

Getting to know the physical reality of a small catchment is often possible. In cases where full access cannot be gained, a good deal can be learned from mapping and aerial photography. Much can be learned from those with long-term knowledge of the particular area, although the information may not be easy to unlock and memories may be selective. Citizen photography is an increasingly valuable resource.

Research into flood estimation on small catchments is being carried out within Environment Agency project SC090031, Estimating flood peaks and hydrographs for small catchments (Phase 2). Some types of local data have been explored within that project, as noted below.

### **3.4.3 Why the use of local data matters**

The effective use of local data – with statistical orthodoxy reinforced by hydrological reasoning – can be expected to lead to more effective spending on flood risk management and to better planning decisions. In some cases, it may save lives.

If a flood leads to major damage or loss of life in a situation where unsuitable development has had a major hand in the formation or routing of the flood or in the extent of impact on people and property, the method of flood estimation is likely to come under close scrutiny. Historical precedent may reveal that the particular stream or site is uncommonly flood-prone. Such potential is sometimes revealed by detailed study of local data. It is seldom revealed in standard assessments.

The use of local data is about more than the refinement of QMED or time to peak. The authors of this report are of the opinion that practitioners should be encouraged to take responsibility in important or difficult cases by being given the freedom to apply local data. They also recommend that clients who commission flood studies should ensure

that budget and timescales allow for, and technical specifications require, consideration of local data appropriate to the scale and importance of the project.

Unless there is scope for the hydrologist to develop and apply reasoned judgement, some of the art of flood frequency estimation will be lost. By promoting the effective use of a wider range of information, FEH Local will help to build (or at least maintain) a pool of experience from which to monitor and question the click-button approach to flood estimation.

## 3.5 UK guidance on the use of local data

Some uses of local data in UK flood estimation have been promoted relatively widely over several decades. The Flood Studies Report (NERC 1975) provides a technique for incorporating historical data into flood frequency analysis, while the FEH exhorts readers to refine generalised estimates of flood frequency by reference to local data (Institute of Hydrology 1999).

Important early references are the specific suggestions made in Flood Studies Supplementary Report No. 13 (Institute of Hydrology 1983) and the case studies presented by Reed (1987).

While the use of local data was embedded within the FEH philosophy, the availability of digital catchment data led some to expect that their use would become automated. This expectation was one of a number of factors – including the demand for generic maps of fluvial flood risk – that led to the incorporation of local data typically becoming more prescriptive. Reed (2002) reviews ways in which flood risk estimates might be strengthened in the digital age.

## 3.6 River flow and water levels

FEH methods make extensive use of flood peak data and, in the case of ReFH, flood hydrograph data. Other ways in which measurements of river flow or level may be helpful in the estimation of design floods are discussed below.

### 3.6.1 Peaks-over-threshold data

POT data series are those that contain the magnitude of every flood peak above a given threshold. In contrast to annual maximum data, some years may contain no peaks and others many. POT series therefore contain more accurate information on the distribution of large flood events and, depending on the threshold value, may contain more information (flood peaks) in total.

The NRFA generally sets the threshold in any given catchment to include an average of 5 events per year<sup>3</sup>, providing 5 times more data than an annual maximum series. This extra information is especially useful for short records; the FEH recommends using POT data to estimate QMED in records shorter than 13 years (Institute of Hydrology 1999). Further, the consideration of peaks, rather than years, also means that it becomes possible to estimate the return periods of events that typically occur more than once per year.

Given the generally larger number of data points available in a POT sample compared with the annual maximum sample available at the same station, the use of peaks-over-threshold data is likely to reduce the uncertainty in flood estimation. Indeed, in statistical terms, a larger sample can only deliver more information, as shown

---

<sup>3</sup> <http://nrfa.ceh.ac.uk/peaks-over-threshold>

empirically by Bezak et al. (2014). Furthermore, there are theoretical results showing that using POT for flood frequency analysis asymptotically delivers the same inference obtained when using annual maximum (see Madsen et al. 1997).

No use of POT data was made in the development of the improved FEH statistical methods described in Environment Agency (2008). The report points out that the POT data were not readily usable without further quality controls and additional work. The original FEH statistical model instead recommended the use of POT data for the estimation of QMED in stations for which only short records are available. This could be still advantageous, compared with using the QMED regression (see Section 2.4.2), although no comparisons of the estimated QMED and uncertainties have been carried out. However, POT extraction is unsuited to catchments with high baseflow, which restricts the value of POT data in developing generalised methods suitable for all catchment types.

At present, flood frequency estimation using POT series is not as advanced in the UK as flood frequency estimation using annual maximum series, although the use of POT series could be beneficial, in particular when only few years of data are available at one station.

### 3.6.2 Gauged baseflow index

Baseflow index (BFI) quantifies the proportion of long-term mean streamflow at a site that derives from baseflow. BFI was developed in the Low Flow Studies report (Institute of Hydrology 1980) from an idea by Lvovitch (1972). In the UK, the NRFA holds gauged BFI values for approximately 1,500 catchments, which it has published in the UK Hydrometric Register (Marsh and Hannaford 2008). The method used to derive the BFI values found in the UK Hydrometric Register is well defined by Gustard et al. (1992) and hence equivalent values can be calculated from any streamflow record of sufficient length and quality. It is possible to calculate BFI at some gauges where the measurement of high flows is not sufficiently accurate to enable extraction of annual maximum flows from which QMED could be estimated directly.

In the FEH, the BFIHOST catchment descriptor provides a measure of the baseflow index estimated from the HOST classification of UK soils (Boorman et al. 1995). BFIHOST is one of the catchment descriptors used in the QMED equation for ungauged sites and, in the absence of gauged streamflow calibration data, the estimation of 5 of the 6 ReFH model parameters and initial conditions depend directly or indirectly on BFIHOST.

Locally gauged BFI should be considered more representative of the catchment in question, as the HOST dataset is derived from low resolution 1:250,000 maps and the same BFIHOST value can be given by many different combinations of soil types. Although not without limitation, the integration of gauged estimates of BFI with indirect estimates based on soil mapping (and other descriptors) is well illustrated in Chapter 5 of Volume IV of the Irish Flood Studies Update (Mills et al. 2014).

In a UK context, the ReFH and FEH statistical methods are built and calibrated around the use of BFIHOST. It is therefore expected that additional calibration would be required to minimise their *fse* values with respect to gauged BFI.

A test carried out on 528 gauged UK catchments suitable for pooling showed that, on average, substituting gauged BFI for BFIHOST results in a greater error in the estimate of QMED. The mean logarithmic error in  $QMED_{BFIHOST}$  was 0.006 compared with an error of -0.070 in  $QMED_{BFI}$ . At 52% of gauges, using gauged BFI yielded a poorer estimate of QMED.

However, the 5 largest errors in the estimate of QMED (which were all overestimates) were all reduced when gauged BFI was substituted; 4 out of the 5 were on catchments

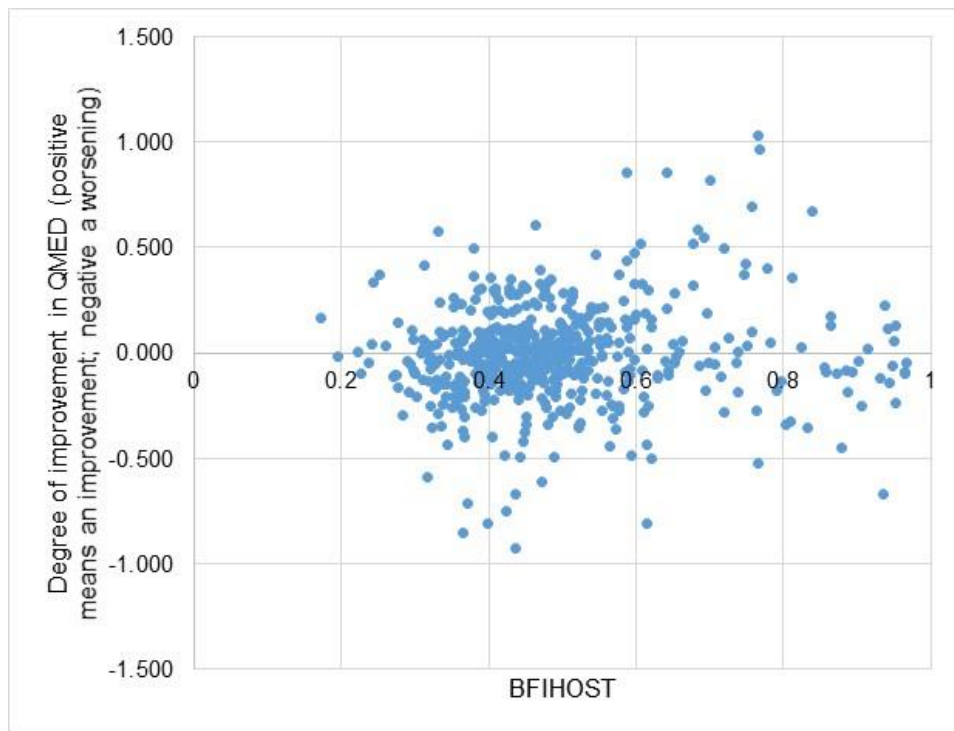
with high BFIHOST (>0.65) and the fifth was on a karst catchment with some drift cover.

As can be seen in Figure 3.1, no significant relationship was found between the degree of improvement (or worsening) in the estimate of QMED and the value of BFIHOST. The degree of improvement plotted on the y axis is defined as:

$$|\ln(QMED_{BFIHOST}) - \ln(QMED_{obs})| - |\ln(QMED_{BFI}) - \ln(QMED_{obs})| \quad (\text{equation 3.1})$$

where  $QMED_{obs}$  denotes the estimate of QMED obtained directly from annual maximum flows.

The implication of these findings is that use of gauged BFI in place of BFIHOST cannot be expected to lead to an improvement in QMED. Although there are a few catchments for which this would reduce large errors, there is no straightforward way of identifying these catchments. However, it could be worth comparing  $QMED_{BFI}$  with  $QMED_{BFIHOST}$  and investigating the catchment or flow data further when they are very different. In reality, this situation will occur only occasionally because, at many gauges where BFI can be calculated, it will also be possible to estimate QMED directly from annual maximum flows.



**Figure 3.1 Improvement in the estimate of QMED as a result of substituting gauged BFI for BFIHOST**

### 3.6.3 Flow gauges with uncertain flood flow measurements

Some flood studies ignore flow gauges that are not thought to give reliable measurements of flood flows. Others argue that, particularly on unusual catchments, having some data is better than having no data. These different perspectives can lead to large discrepancies in the results of studies. Two projects carried out for the Environment Agency on a major river in the east of England took these different approaches and, largely as a result, the estimated 100-year flows differed by 40%.

Section 3.7 summarises how flow data from gauges with provisional flood ratings have contributed to estimation of design flows throughout the county of Devon.

The practitioner guidance includes some advice on the need to consider gauges beyond those included in the NRFA flood peak dataset.

There is also potential for using information on low to average flows when estimating flood flows. One example is via calculation of BFI as discussed above.

### **3.6.4 Water level gauges**

Water level data can be used, in conjunction with catchment rainfall, to estimate catchment lag time. Volume 4 of the FEH describes how the lag can be converted to an approximate estimate of the time to peak of the unit hydrograph for the FEH rainfall-runoff method. Although publications describing the ReFH method do not mention the possibility of lag analysis, it is sometimes carried out in practice where water level records are available to improve on the estimation of time to peak in the ReFH method.

Time series of peak water levels can be used to estimate the median annual maximum level (LMED) for comparison with the results of a hydraulic model run for the 2-year return period flow.

Other ways in which water level data can be exploited in flood estimation are described in Section 3.6.7 (the Devon Hydrology Strategy) and in the practitioner guidance.

### **3.6.5 Temporary river gauges**

Even very short records of river flow or level can make a significant difference in the estimation of design flows. In a paper entitled 'Gauging the ungauged basin', Siebert and Beven (2009) describe how a few runoff measurements can contain much of the information content of continuous runoff time series. However, they acknowledge the risk of sampling a period that is unrepresentative of the long-term record.

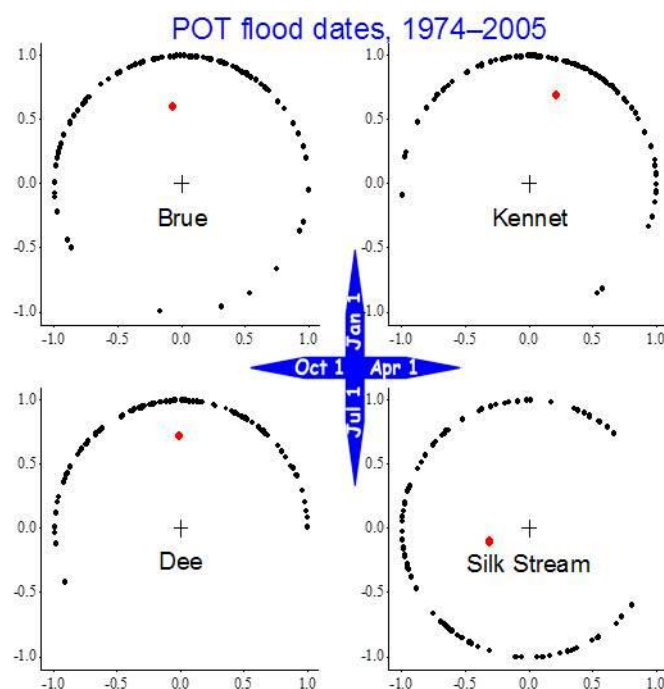
To take an example from the UK, a few months after temporary flow gauges were installed on small watercourses in Bentley, south Yorkshire, one of them recorded a flow that was greater than twice the value of QMED estimated from catchment descriptors. This was a convincing illustration of the highly unusual flood hydrology of the permeable and urbanised catchment, and the short flow records were used to calibrate a rainfall-runoff model that enabled flood estimation by continuous simulation.

In other situations, records of only a few months of flow data manage to capture enough flood events to enable estimation of the parameters of the ReFH rainfall-runoff model, hence improving on a design flow estimated solely from catchment descriptors. Such records may also enable a water balance calculation – a valuable aid to the understanding of catchment processes.

### **3.6.6 Flood seasonality**

The seasonality of flooding is an additional source of information that is not routinely accounted for in most flood estimation studies. WinFAP-FEH software supports the use of flood date information in the diagnosis of pooling group heterogeneity. Several topics in Reed (2011) – most notably circular diagrams (see Figure 3.2), daily mean flow, disparate catchments and surrogate – further promote the use of such information. There is potential for greater use of readily available information on flood dates. Flood dates are generally known for all stations, whether rated for flood measurement or not. Ideally, they are based on POT data.

Plotting conventions sometimes differ. The red dots in Figure 3.2 mark the mean flood days and the degrees of seasonal concentration. These can be used as summary statistics. Non-parametric methods can be used to test the significance of catchment differences in flood seasonality. Fisher (1993) provides a useful reference.



**Figure 3.2 Use of circular diagrams to summarise and distinguish flood regimes**

Source: Reed (2011)

The average seasonality of floods inferred from POT series is not always a reliable guide to the seasonality of the very largest floods. This is because the very largest rainfalls are often a summer phenomenon, and so it is quite common to see the majority of floods taking place in winter but the largest floods occurring at other times of year.

An additional index based on POT flood dates is the coefficient of variation of recurrence intervals (CVRI). As the name indicates, it is defined as the standard deviation of the intervals between floods (conveniently measured in days) divided by the mean of those intervals. It is a measure of the (temporal) irregularity of flood occurrences.

In districts where soil moisture deficits (SMDs) are seldom large, flood occurrences tend to be more regular and (by implication) to be more directly related to the pattern of heavy rainfall occurrences (Bayliss and Jones 1993). This gives rise to a small value of CVRI (for example, in the range 0.8 to 1.2). In contrast, in districts where SMDs are often large, flood occurrences are more likely to occur in batches, with longer flood-free periods in between. This gives rise to much higher values of CVRI (for example, in the range 1.8 to 2.2).

The CVRI research showed some promise but was side-lined in the FEH project, where seasonal indices were judged easier to comprehend. Moreover, trials revealed CVRI to be quite sensitive to the period of record available. A rather complicated

procedure was devised to stabilise values prior to mapping flood irregularity (see page 180 of FEH Volume 3).

The irregularity of flood occurrences says something about the rainfall-runoff behaviour of the catchment. When CVRI is large ( $>1.7$  might be a useful criterion), the correspondence between heavy rainfall events and flood events can be especially hard to decipher. In such cases, some analysts may be reluctant to give weight to flood estimates based on a design event method. One example is the Harpers Brook catchment near Corby (Station 32003). Reed (2011) notes that the 100 largest floods in the 25 years commencing 1 October 1961 had a mean flood date of 18 February, whereas the 100 largest daily rainfalls in the same period had a mean date of 15 August. The seasons of river flooding and maximum one-day rainfall are diametrically opposed.

### 3.6.7 Hydrograph width analysis

Flood estimation research and practice have progressed down somewhat different channels in Ireland. New methods were published in 2014 as the Flood Studies Update (FSU), with digital methods implemented through the FSU web portal (<http://opw.hydronet.com>).

A distinctive feature is that design flood hydrographs are based on 'hydrograph width analysis' and constructed without recourse to a rainfall-runoff method. The approach loosely follows that put forward by Archer et al. (2000) and has seen application on several UK rivers.

Research at the National University of Ireland Galway on hydrograph widths suggests that valuable information on flood response times might be gained from the study of the rising limbs of hydrographs alone. Such studies are relevant to flood forecasting as well as to the construction of design flood hydrographs. There is scope to apply such methods to records from level-only river gauges.

## 3.7 Example: The Devon Hydrology Strategy

The Devon Hydrology Strategy was an initiative to develop consistent and believable estimates of design flood flows along 3,000km of watercourses in the county. It incorporates the wider use of local data in a county-wide approach to flood risk estimation. The initial study (Royal Haskoning 2007) exploited flood peak data from 82 gauging stations in Devon, many more than the 29 included in the NRFA peak flow dataset. Fifty-nine of the additional stations provided estimates of peak flows through the use of provisional flood ratings.

The stations used and their flood ratings are kept under review and design flood flows are periodically recalculated. The 2012 update by Royal Haskoning (2013) focused chiefly on 65 stations.

The Devon Hydrology Strategy is important because it questions how the middle ground between main river and small catchment applications is dealt with in basin-wide flood risk mapping. In the Devon Hydrology Strategy, the Environment Agency's preference for spatially consistent estimates is met by invoking regional relationships between the specific design flood flow (that is, expressed in  $\text{m}^3/\text{s}$  per  $\text{km}^2$ ) and catchment area. A neat feature of this approach is that flood estimates are developed first at gauged sites across the network and then ratcheted out to nearby ungauged sites. Special checks are made to resolve mismatches at confluences.

The primary role accorded to catchment size is likely to downplay the influence on flood flows of catchment heterogeneity in soils, topography, storage and land use. Thus, the

approach may work less well in river basins of mixed character. The method is understood to heighten the role played by pooled analysis of flood data but to diminish that played by single-site analysis.

The use of fixed regions in the Devon Hydrology Strategy may lead to flood growth rates applied to broadly similar small catchments differing only because the catchments lie on opposite sides of a regional divide. This was a recurrent criticism of the Flood Studies Report (NERC, 1975) regional growth curve method and led the FEH research team to develop a 'region of influence' (Burn 1990) style of pooling. While the greater use of local flood data within the Devon Hydrology Strategy is to be commended, a better approach might apply local data in a manner that modifies estimates (chiefly for large and medium-sized catchments) close to the sites providing additional data, without unduly changing estimates for small ungauged catchments remote from the sites providing additional data. However, such a balance will not be easy to achieve.

## 3.8 Climate

### 3.8.1 Climate variables

Some features of the climate (such as SAAR) are accounted for in UK methods of flood estimation. Others have been considered in research and found to have little power in explaining variation in design flood statistics at a national scale. For example, the steepness of rainfall growth curves was indexed as a new catchment descriptor during the development of the revised FEH statistical method (Environment Agency 2008) but not included in the final recommended methods as it did not have a significant influence when the peak flow dataset was considered as a whole.

There may be potential for local allowances for climatic features, the influence of which may be difficult to formalise in a general way. For example, some users of the FEH statistical method may judge that information on the gradient of rainfall growth curves is a useful factor in guiding the composition of pooling groups for some types of catchment or in some areas of the country, despite the findings of research at a national scale. Overseas methods, including the recent revision of Australian Rainfall and Runoff (Engineers Australia 2016), use rainfall frequency information in the statistical estimation of the parameters for the flood frequency distribution on ungauged catchments.

An example of the striking influence of local climate on specific flood discharges is given by Merz and Blöschl (2008), where neighbouring Alpine catchments with similar area, mean annual precipitation, soils, geology and elevation have specific discharges that differ by a factor of 2. The explanation offered by the authors is that storm tracks tend to follow a particular direction, resulting in orographic influences on precipitation that differ greatly between catchments.

The mapping of the FEH index rainfall RMED took account of:

- local topography (indexed by ELEV10 and OBSTW)
- maritime influence (indexed by SEAWSW)
- the 'continentality' of the climate (indexed by DLILLE)

These descriptors denote the average elevation in the 10km grid square centred on the grid point, the weighted average angle of topographical obstruction to the west, the distance from the sea to the west-south-west and the distance from Lille (in northern France), respectively. Further details are given in Section 7.2 of FEH Volume 2 and in Prudhomme and Reed (1998).



### **3.8.2 Local rainfall analysis**

Except at very long durations, rainfall frequency estimation is a much better resolved problem than is flood frequency estimation. This, in part, reflects the denser network and the longer records available for rainfall. However, the principal reason is that design rainfall depth estimates can be expected to vary relatively smoothly in space once topographic and maritime influences have been accounted for. For ungauged sites at least, this makes:

- estimates of the index rainfall RMED much more reliable than estimates of the index flood QMED
- pooled estimates of rainfall growth much better defined than pooled estimates of flood growth

### **3.8.3 Rainfall accumulations over very long durations**

Some studies of sewer flooding in summer 2012 and of river and groundwater flooding in winter 2013 to 2014 considered frequency evaluations for long duration rainfalls. It is important, however, to recognise that applications of extreme value theory require a large number of nearly independent peaks from which the extreme one is selected. In practice, annual maximum analysis of rainfall depths applies acceptably out to 8-day rainfall depths and (at a very considerable pinch) to 32-day depths.

The annual maximum approach is inapplicable to 60-day or 90-day rainfall depths. In many years, the maximum 90-day rainfall will not represent an extreme event, and the second or third ranking 90-day event may represent a period of below average rainfall. The requirements for extreme value analysis are not met.

One way of making progress is to consider 5-year maximum (rather than annual maximum) 90-day rainfall events. However, exceptionally long rainfall records are then required to estimate event rarity with any confidence.

Similar limitations broadly apply when assessing the rarity of other long duration phenomena, including groundwater flooding.

### **3.8.4 Dealing with climate variability**

Climate variability is an important factor to consider when undertaking specific analyses of flood frequency. The need arises from the flood poor and flood rich periods that are a natural feature of the UK climate. Adjustments are therefore appropriate when estimating QMED at stations with short records. Chapter 20 of FEH Volume 3 considers the problem in considerable detail and presents a sophisticated procedure for adjusting QMED by reference to one or more stations with longer records.

Approaches to adjusting growth curve analysis are less well developed. One approach is to use the original flood data (unadjusted) in single-site analysis of flood growth. Where site records are of short duration, the growth factors contributing to the adopted flood growth curve will be dominated by those from the pooled analysis. The important practical consideration is therefore to adjust the QMED value.

Exploring the sensitivity of pooled growth curves to climate variability is too specialised for general guidance. Academic research tends to concentrate on the sensitivity of single-site and pooled analyses to non-stationary effects, whether these arise from climate variability, climate change or land use change. Valuable references are O'Brien and Burn (2014) and Prosdocimi et al. (2015).

## 3.9 Groundwater levels

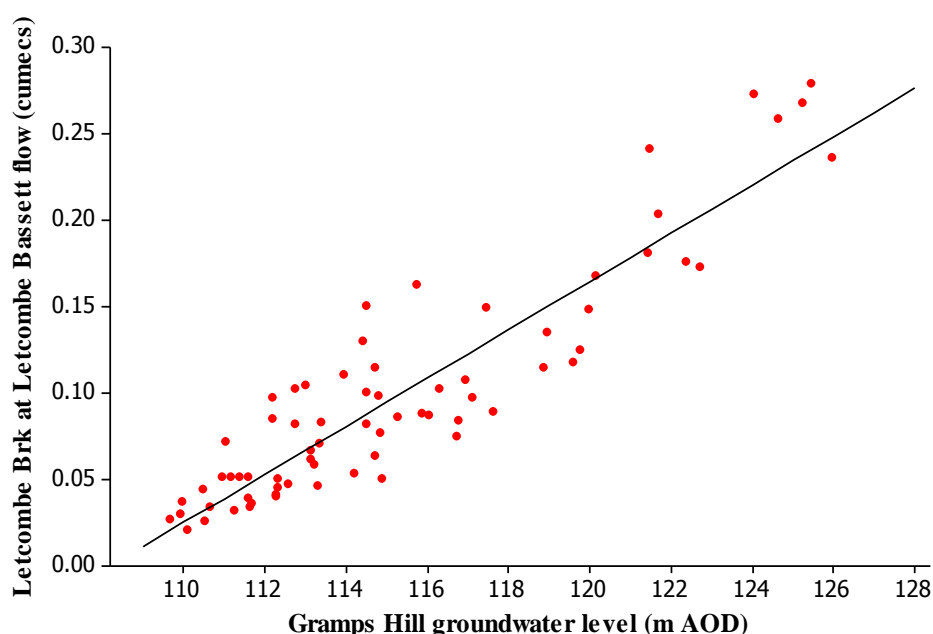
### 3.9.1 Context

A sequence of extreme events – principally since 2000 – has led to groundwater flooding attracting greater attention, especially in southern and eastern England. However, consideration of groundwater flooding lies outside the main scope of this project. The sections below consider the potential for incorporating local information on groundwater when estimating fluvial flood frequency.

### 3.9.2 Relationship between groundwater levels and river flows

On a highly permeable catchment, there may be a strong association between river flows and groundwater levels in the vicinity. Figure 3.3 is based on 70 days (earliest date 29 February 2000, latest 27 September 2007) for which both Letcombe Brook flows and Gramps Hill groundwater levels were available. Gramps Hill is in the catchment periphery, on the scarp slope of the Berkshire Downs.

The flow data at Letcombe Bassett make use of the Gramps Hill groundwater level irrelevant. However, the example illustrates the scope for proxy use of groundwater level data in certain situations where the record is long and relatively complete.



**Figure 3.3** Link between Letcombe Bassett daily mean flow and Gramps Hill groundwater level

Source: Reed (2008)

### 3.9.3 Changes in groundwater abstraction

It is well known that late 20th century reductions in groundwater abstraction led to rising groundwater levels in the aquifer beneath London and that this posed a flood threat to underground infrastructure. This groundwater resource and liability is subject to close scrutiny (see, for example, Environment Agency 2014).

Elsewhere, the link between groundwater abstraction and flood frequency is little noted. Reed (2011) reports a case in the Cotswolds where Thames Water was obligated to make ‘sustainability reductions’ in abstractions from groundwater in order to reduce the frequency with which the Ampney Brook dried up, without regard to whether this might have implications for flood frequency to development accustomed to the abstraction. Reed (2014) reports a case in West Yorkshire where cessation (after 115 years) of groundwater abstraction to supply a large hospital appears to have aggravated flooding problems in the village that expanded to serve the institution.

These examples illustrate that changes in groundwater abstraction can have implications for flood management where the aquifer is relatively shallow and responsive.

### 3.10 Recent floods

In ungauged catchments or those with flow records that are too short or intermittent to permit statistical analysis, it may be possible to acquire information on recent flood events that is helpful in guiding the estimation of design floods. Videos, photographs and contemporary maps of flooded extents can provide valuable feedback on whether watercourses and floodplains are behaving largely as modelled.

Post-flood surveys can estimate peak discharges, providing valuable evidence, in particular for flash floods (Rosbjerg et al. 2013). The resulting estimates can be used to construct flood envelope curves, that is, plots of specific peak discharge against catchment area, although these have only limited use in flood frequency estimation because they describe only the magnitude of floods and not their frequency.

A feature of some flood modelling studies is that, after the estimated design flood is applied to a hydraulic model, the resulting flood outline is thought to be inconsistent with local information. For example, it may show that an area known to have flooded is outside the modelled flood outline for a long return period event, such as the 100-year or even the 1,000-year flood. In such cases, it is necessary to ensure that the reported flooding was due to fluvial sources and to examine whether it may have been exacerbated by hydraulic effects – such as blockage of structures by debris – that may not be represented in the hydraulic model. It is also necessary to judge whether the discrepancy is likely to be due to errors in the design flood estimate or in the hydraulic model.

Several examples of such studies were examined during the research. In nearly all cases, flows estimated from FEH methods were thought to be underestimated in the light of evidence from observed floods. Most but not all were on ungauged catchments. Approaches to resolving the apparent discrepancies included:

- adopting an upper confidence limit from FEH methods as the preferred design flow estimate
- reducing time to peak in the ReFH method by estimating lag only from more intense rainfall events to allow for a reduction in lag time with rainfall intensity
- modelling of pluvial flood extents to assess whether previous floods may have been due to direct rainfall rather than fluvial events
- making improvements to hydraulic models

The practitioner guidance developed in parallel with this report presents a structured way for analysts to weigh up evidence of the impacts of recent floods and decide whether or not to adjust flood frequency estimates as a consequence.

## 3.11 Historical floods

### 3.11.1 Methods

There is considerable experience in the collection and use of historical flood data, in addition to specialist research by, among others, Macdonald and Black (2010), Macdonald (2014) and Macdonald et al. (2014). A major review of the use of historical data in 15 European countries plus Turkey and the guidelines currently in use in 9 of these is given by Kjeldsen et al. (2014c)

The utility of including historical information in flood frequency analysis has long been investigated (for example, Benson 1950, Leese 1973). Extensive practical guidance was provided by Bayliss and Reed (2001), who favour graphical methods over the use of formal statistical methods for combining gauged and historical flood data.

There is a large body of literature on statistical methods for incorporating historical data in flood frequency estimation. These can largely be divided into methods that extend the L-moment approach of Hosking (1990) and methods based on maximum likelihood.

Wang (1990a) introduced partial probability weighted moments (PPWM), which modify probability weighted moments (PWM) calculations to allow the inclusion of right-censored data (that is, data for which the event magnitude is known only when larger than a given perception threshold). Wang (1990b) showed that the use of PPWM could improve the estimation of high return period events in cases in which some information would be available on large historical floods, in particular for negatively skewed GEV distributions. Other approaches that modify L-moment type calculations to include censored data include those described by Hosking (1995), the expected moments algorithm presented by Cohn et al. (1997) and the expected probability weighted moment estimator presented by Jong-June et al. (2011).

Likelihood-based methods can easily accommodate historical flood peak data and have long been used (Stedinger and Cohn 1986). Once the likelihood is defined, maximisation is possible, either directly through numerical methods or through Bayesian methods, as detailed, for example, by Reis and Stedinger (2005) and Gaume et al. (2010). Likelihoods can be specified even if all that is known is that a given number of threshold exceedances have occurred, without knowledge of the magnitude of the historic events (Stedinger and Cohn 1986). Furthermore, maximum likelihood methods have been employed to include data that exceeded different perception thresholds, or for which (rather than a point estimate of the flood magnitude) an interval estimation was available (Gaál et al. 2010). Neppel et al. (2010) used the Bayesian framework to include uncertainties around both the rating curve and the historical event peaks in the estimation of flood curves. Their method addresses the fact that it is not only the historical events recorded outside the systematic record that are affected by some degree of uncertainty around the peak value.

According to Kjeldsen et al. (2014c), there seems to be a wider adoption of likelihood-based methods in EU countries, mostly owing to the greater flexibility they allow in terms of the type of data that can be included in the analysis. The proposed Bayesian likelihood-based methods, which have been developed largely by the academic community, are quite complicated. They also require the user to provide information about a number of uncertainty elements and to be able to assess the goodness of the performance of the FEH statistical method, making it unlikely that these models can be easily adopted by practitioners. Spain appears to be the only EU country in which the use of PPWM methods is recommended, while the expected moments algorithm is still used in the USA. The Flood Studies Update web portal in Ireland envisages that L-moment methods for incorporating historical flood peaks will be developed further.

The FEH Local research has investigated both the likelihood and L-moment methods for incorporating historical data (see Section 0). There is experience in applying likelihood methods in UK practice, including on 7 catchments in Northumbria (JBA Consulting 2008) and in a lowland small catchment in southern England (Macdonald et al. 2014).

A combination of graphical and analytical approaches is likely to be preferable when integrating gauged and historical flood data. Ideally, the analytical approach should strike a good balance between the complexity of the statistical model and the ability to accommodate a number of types of local data.

### **3.11.2 Use of historical flood data**

Historical information is seldom exploited to its fullest extent in UK practice despite the frequent emphasis on the importance of flood history in the academic literature, in the Environment Agency's Flood Estimation Guidelines<sup>4</sup> and on training courses. Possible reasons include:

- the complex computational requirements of some methods (as suggested by Kjeldsen et al. 2014c)
- expectations that flood estimates can be obtained from computer software alone
- inadequate training or mentoring of hydrologists
- unclear guidance (mentioned by Clark 2014)
- restrictions imposed by project budgets or timescales
- a lack of specific requirement for historical review in project briefs

When historical information is exploited, numerous studies have shown substantial changes in the estimated design flows. For example, Black and Fadipe (2009) showed that estimated 100-year flood flows at 3 out of 4 sites increased by more than 50% as a result of incorporating reliable historical information. JBA Consulting (2009) demonstrated increases of up to 54% in the estimated 100-year flow. Clark (2004) used historical flows on the River Till in Wiltshire to estimate flows several times larger than the results of FEH methods.

Conversely, Macdonald and Black (2010) demonstrated that the FEH estimates of 100-year flow at York were implausibly high, as the estimated flow rates had not been reached in the entire 737-year historic series. This is because, at very large flows, flood peaks arising from the Nidd, Ure and Swale are attenuated by widespread inundation of floodplains upstream of York.

Hydraulic models are occasionally used to infer flows from levels when interpreting historical flood information. In lowland Britain, in particular, river channels and their floodplains have been changed by many factors, including agricultural improvements, the encroachment of development, flood alleviation works, weir alterations, bridge alterations and dredging; this is in addition to natural effects related to alluvial deposition and (occasionally) channel migration. However, there are often good documentary records of 19th century and 20th century floods, the latter typically supported by photographs.

---

<sup>4</sup> internal document for Environment Agency staff and consultants working for the Environment Agency

There is sometimes a reluctance to acknowledge flood marks from exceptional events, such as the Tame flood of 13 July 1872 at Uppermill in the south Pennines. A typical reason given is a lack of confidence in river and floodplain conditions at the time of the flood. An unwillingness to research exceptional events can lead to one-sidedness in the use of historical flood data. This is troubling, since even some very rough information on the highest flood event above a threshold can significantly change the estimated flood curve. Payraastre et al. (2011) show that augmenting relatively short systematic records with information on historical peaks reduces the uncertainty around the estimated flood curve, even if the added information consists of just few very rare events, for which potentially only the perception threshold is known.

### 3.11.3 Chronology of British Hydrological Events

The Chronology of British Hydrological Events (CBHE) ([www.cbhe.hydrology.org.uk](http://www.cbhe.hydrology.org.uk)) is an exceedingly valuable resource, largely set up and maintained by enthusiasts (Black and Law 2004). A relatively small number of professional hydrologists have contributed most of the information.

Its strengths include the dual indexing by river basin and date, and the ability to search for specific text. One weakness is the episodic (that is, event-by-event) nature of the catalogue. In some cases, reports of a large flood refer back to earlier events. More typically, the user is left uncertain of the length of the historical flood series within which the noted flood is thought to be (say) the largest or second largest.

The CBHE was relaunched in 2016. The way in which people access and share information had changed greatly since the original website was developed. The relaunched version includes a map interface, although no previous entries in the database were georeferenced. Contributors are encouraged to add map references for new entries. There is a need to classify the character and likely quality of information held and to encourage user expansion of the catalogue.

There is potential for combining with a current initiative by the Environment Agency to develop a geographical information system (GIS) interface to a new chronology of flash flood records in the north-east, north-west and south-west of England compiled as part of the SINATRA project (Archer and Fowler 2015).

A new way of sharing and accessing local data, including historical floods, is proposed in Section 4.4 of this report.

### 3.11.4 Other sources of historical data

There are many potential sources of information about historical floods, including some deriving from predecessor organisations to the Environment Agency.

Online access to existing reports that summarise historical flood information for a particular river might be useful, even though specialists argue that there is no substitute for the user consulting the source document rather than a derivative account. As pointed out by Bayliss and Reed (2001), it is necessary to judge whether the originator of historical information was well positioned to report the event accurately and impartially.

A proportion of CBHE entries derive from British Rainfall yearbooks. The Met Office has made all the yearbooks available online<sup>5</sup> and the ability to find references to

---

<sup>5</sup> [www.metoffice.gov.uk/learning/library/archive-hidden-treasures/british-rainfall](http://www.metoffice.gov.uk/learning/library/archive-hidden-treasures/british-rainfall)

particular places, rivers or raingauges is relatively good. Understandably, searches are occasionally constrained by mistakes in optical character recognition.

The UK benefits from extensive archive, library and newspaper records. As further historical sources are summarised and/or loaded onto the web, the opportunities for finding forgotten floods widens.

One use of the CBHE (and other sources) can be to provide a candidate list of possible flood dates on which to base a targeted search of local newspapers. Online reference to, and searching of, local newspapers is a very considerable resource. Nevertheless, the local library can still provide information that is readily available nowhere else.

## 3.12 Palaeofloods

Palaeoflood analysis is the study of old floods. The US Bureau of Reclamation distinguishes 2 broad categories of palaeoflood (also spelt paleoflood) data, one based on fluvial geomorphic evidence and the other on botanical evidence. Essentially, palaeoflood data are distinguished from historical flood data by the absence of human observation of the flood.

Appendix A contains an in-depth independent review of palaeoflood data and their potential for use in flood risk assessment in the UK. It concludes that traditional flood frequency analysis based on instrumental flow records in the UK needs to be radically rethought so as to protect lives, property and nationally important energy and transport infrastructure assets.

Palaeohydrology is an extensive field of research; a review of progress in western Europe is provided by Brown (2003). Palaeoflood research is an active topic, which has found application in virgin terrain to assist in setting design floods for major reservoirs such as those owned and operated by the US Bureau of Reclamation. A window on this kind of application is provided by O'Connell et al. (1998).

Palaeoflood studies have been conducted in the UK, most notably in some upland catchments of the northern and western UK. A report from the Lloyd's Emerging Risks Team notes evidence from the deposition of boulders and large cobbles transported in extreme floods (Lloyd's 2012). When a boulder comes to rest after a flood, lichens are able to colonise the newly exposed rock surfaces within a few years. These can be dated through lichenometry, which is the measurement of lichens to estimate the age of the deposits on which they grow. Their growth rates can be used to estimate a minimum age for the flood event. By way of calibration, lichen growth rates over the last 200–400 years have been measured on rock surfaces of known age, for example, gravestones and built structures. The Lloyd's report notes that this method has created records of extreme flood events in parts of the northern Pennines, Lake District, Yorkshire Dales and Brecon Beacons, going back to 1750 and even to the 17th century in some areas. In such studies, it may be possible to estimate the minimum flood flow consistent with movement of a boulder of given size, shape and mass. Yeo (2010) applied the same method at Lynmouth in Devon to estimate the discharge of the 1952 and other floods.

An assessment of 21st century floods in the UK uplands by Foulds and Macklin (2016) using lichen dating showed that recent floods have not been unprecedented in terms of their frequency or magnitudes.

Until recently, palaeoflood studies have generally been less productive in lowland Britain. One reason for this is the agricultural use of riparian land. Lowland agriculture typically runs close to the river bank and modifies upper soil horizons through ploughing, embanking and other activities. In the heavily populated environment of

lowland Britain, it may be difficult to find palaeoflood evidence that is assuredly undisturbed by man.

However, some authors offer a more upbeat assessment of the potential for palaeoflood studies in lowland Britain. Jones et al. (2010) described a technique for the analysis of floodplain sediment sequences which have been found to maintain their depositional integrity at some sites. Vertical profiles of elemental composition in sediment cores can be obtained using X-ray fluorescence, and the ratios of concentration of various elements used as a proxy for grain size and thereby for flood magnitude. Ages can be assigned to the flood record using techniques including carbon-14 dating. Several examples of studies in lowland Britain are summarised below.

A study by Paine et al. (2002) had some success in reconstructing historic floods by analysing sediments from embanked flood plains. The study was chiefly based on caesium-137 analysis and focused on identifying sediments from embankment breaches on the River Tay at Bloody Inches over the period 1954 to 2000. The research concluded:

‘At sites where flood plain geomorphology offers suitable preservation conditions, the general applicability of this approach has the potential to elucidate flood histories and refine risk estimates in the absence of long-term instrumented hydrological records’.

Jones et al. (2012) described the production of a 3,750-year-long flood record on the floodplain of the upper Severn at an altitude of 60m using X-ray fluorescence. The historical flood of 1795 AD appears to have been the largest since around 200 AD. The authors acknowledged that the method did not currently provide a means of determining the discharges of floods. They also stated that the range of channel and floodplain types suitable for the application of this method of flood reconstruction remained to be established, although the presence of areas of uninterrupted fine-grained sediment deposition was a prerequisite, as was a means of dating the sediment sequence. The availability of a long instrumental or historical flood record for comparison with the sedimentary record will increase the information that can be derived from it.

It may be that the main value of palaeoflood investigations in a UK context is as a precautionary step in the estimation of extreme floods for the protection of major infrastructure such as reservoirs or nuclear installations. Evidence of past extreme events may help guide estimation of the 10,000-year flood, for example.

Consideration of palaeoflood data is advocated by Bayliss and Reed (2001) for situations where potential hazards are extreme, but the advice has not always been heeded to date.

## 3.13 Catchment properties

### 3.13.1 Case experience

In UK practice, many catchment properties are already accounted for by the FEH catchment descriptors. However, there are examples in the literature of site-specific hydrological reasoning being applied to allow for the influence of unusual catchment properties. Examples include Merz and Blöschl (2008), who show how a catchment in Austria has a much lower specific discharge than its neighbours despite similarities in rainfall, slope, soils and land use. The anomaly is thought to be due to the presence of



local gravel deposits in the valley floor, which allow for subsurface flows. Note that the example refers to a different pair of catchments to that quoted in Section 3.8.1.

Merz and Blöschl (2008) also gave examples of how field visits or examination of topographic maps can yield clues about the flood frequency behaviour of catchments, such as the degree of incision of valleys, the presence of indicator plants or the characteristics of the river channel. They stated that:

'It would not be possible to predict the differences in catchment response between the two catchments on the basis of the quantitative catchment attributes and formal methods alone. In contrast, soft information obtained through a visual examination of the catchments during site visits may help tremendously. Clearly, site visits are instrumental in a hydrological assessment'.

While many hydrologists would concur with this as an aspiration, site visits are not always budgeted for in UK flood hydrology practice. This is understandable when studies of large geographical areas are commissioned, for which familiarisation with the gauging network is the priority. However, it is a worrying omission if the study is of a small catchment in a sensitive area. Some types of information such as an indication of the channel form, evidence of spillage from neighbouring catchments or anecdotal information from local people can only be properly gained by site inspection. However, there is much that can be inferred about the characteristics of a small catchment by close inspection of detailed maps and aerial photographs. Both are valuable for identifying unusual features such as quarries, bogs and woodland that might be missed in a digital summary of catchment properties.

Information on local soils and geology is widely available within the UK. As well as the national 1:250,000 soil map, some areas have more detailed soil maps, for example, at 1:63,360, 1:50,000 or 1:25,000 scale.

Where development has taken place over many eras, examination of historical maps can be helpful in understanding how the once-natural drainage system has been modified by development. This can be especially informative in areas where springs were formerly a notable feature and/or where the groundwater was later exploited as a local water resource. Understanding the evolution of the catchment is relevant to putting its flood history into perspective.

### **3.13.2 Allowance for unusual catchments in design flood estimation**

Often it is a matter of experience and local knowledge to know how best to exploit detailed information about soils, geology, land cover or land use in the estimation of design flows. Research such as that in Project SC090031<sup>6</sup> has shown that some catchment properties, such as forestry and agricultural land use, have little discernible influence on flood frequency, even at the scale of small catchments.

At a bare minimum, it is essential to confirm that the catchment boundary and soil type given by standard digital representations correspond reasonably to those shown on detailed maps. Where doubt arises, a site inspection should be considered mandatory. The practitioner guidance gives advice on the types of features to look for.

Once the character of the ungauged catchment has been assessed by review, inspection or analysis, the first step is to confirm that estimates of FEH catchment

---

<sup>6</sup> Estimating flood peaks and hydrographs for small catchments.

descriptors reflect reality. Where appropriate, adjust descriptor values and keep an explanation (to allow audit).

Where the unusual catchment property is not explicitly represented in the formal FEH procedure, one approach is to allow it to influence the selection of a donor catchment.

Some sites become subject catchments because of recent or perceived flooding problems. Should evidence be found that the subject catchment responds to an unusually wide range of flood event types (for example, compared with the donor catchment), assessments should recognise that the subject site is unusual and may suffer greater flood risk than indicated by standard assessments.

Other sites enjoy reduced flood risk through their location downstream of an extensive undefended floodplain or other flat area such as a wetland. While this feature is maintained, it is to be expected that flood risk may be smaller than indicated by standard assessments.

Allowances for unusual catchments are inevitably subjective. Those unwilling or unable to accept non-standard methods need to find a way of accommodating the unusual catchment within standard methods. One approach is to accept delay and install instrumentation. Another is to add a higher than normal freeboard allowance (Residual Uncertainty Allowance). An alternative is to seek an independent expert opinion.

Even without any models or software tools that enable practitioners to account for information on additional catchment properties or features when estimating design flows, hydrologists should be capable of applying their judgement when faced with unusual catchments. This requires a grounding in catchment science and a sound understanding of the assumptions and principles of modelling techniques. However, both these raise educational and training implications that could potentially come into conflict with drivers for efficiency or the prescriptive use of analytical tools.

## 3.14 Channel dimensions

### 3.14.1 Background

Estimation of QMED from channel dimensions provides a potential alternative to estimation from catchment descriptors. It is recognised that, in many natural rivers in the UK, the water level in the main channel reaches bankfull every 1–2 years; it may therefore be possible to estimate QMED from channel dimensions.

The FEH provides such a method based on research conducted by Geraldene Wharton in the late 1980s and early 1990s. Her work was carried out in 2 stages. The first, described in Wharton et al. (1989), used existing channel survey data held in the Surface Water Archive. The second, described in Wharton (1992), involved new field surveys to obtain a consistent dataset and to extend the sample of sites for analysis. The dataset described by Wharton (1992) comprised 75 sites with field-surveyed data and 109 with data from archived sources.

The study described by Wharton (1992) selected gauging stations with at least 5 years of good quality flood peak data and no artificial modification of the flow regime or channel. Sites were excluded that had undergone re-sectioning, realignment, bank protection or clearing of vegetation or debris to such an extent that channel dimensions were affected. These restrictions are presumably the reason why there were very few sites in the Midlands or east of England, and hardly any sites on urbanised catchments. They may also explain the absence of large catchments in the dataset.

Channel dimensions obtained were width, cross-sectional area and mean depth, all calculated at a selected reference level, either bankfull or overtopping.

Wharton (1992) analysed the data from surveyed and archive sources separately, because it was not possible to accurately represent bankfull levels from archived cross-section surveys. Regression equations were developed to predict the mean annual flood, the 1.5-year flood and the 5-year flood from 3 dimensions:

- channel width
- cross-sectional area
- ratio of width to mean depth

One set of equations was developed for bankfull dimensions and another for overtopping dimensions.

One conclusion was that equations based on cross-sectional area are superior to those based on channel width, since the former is more representative of the channel size and particularly so in channels with a riffle-pool sequence. This conclusion makes logical sense and it might be expected that a measure of channel slope would be a beneficial addition to the regression since, in Manning's equation, flow is a function of channel cross-sectional area and shape, slope and channel roughness. However, the conclusion does not appear to be borne out by the statistical performance of the equations in Wharton (1992); for example, the regression of mean annual flood on channel width for bankfull stations had  $R^2 = 0.78$ , whereas that on channel area had  $R^2 = 0.73$ .

### **3.14.2 FEH and channel dimensions**

Several years after Wharton's work was published, some of the data were re-analysed at the Institute of Hydrology as part of the FEH research. Bankfull channel width (BCW) data from 65 sites were used to develop a regression for QMED (given as Equation 5.1 in FEH Volume 3). These were taken from the 75 surveyed sites described by Wharton (1992), with some sites apparently rejected owing to lack of QMED data or urbanisation. The fitted equation explained over 80% of the variation in QMED, with  $fse = 1.73$ . However, the FEH does not explain why BCW was chosen in preference to other measures of channel geometry.

This approach to the estimation of QMED has been rarely applied in practice since the publication of the FEH. The motivation for its original development was to provide a rapid way of estimating a preliminary design flow without the time-consuming work necessary to abstract catchment characteristics from maps when applying Flood Studies Report methods. Although this consideration is not applicable when using FEH methods that employ digital catchment descriptors, it is worth considering whether there are some catchment types or situations where estimation of flood flows from channel characteristics may give more reliable results than using standard methods. This question is explored in Section 4.3.

### **3.14.3 Subsequent developments in the literature**

Since the early 1990s, various studies have attempted to develop equations that represent the relationship between hydraulic geometry and discharge. Wharton and Tomlinson (1999) described a similar approach to Wharton (1992) in 4 developing countries – Java, Burundi, Ghana and Tanzania. Regression equations to estimate mean annual flood from active channel width had  $R^2$  values ranging from 0.71 in Ghana to 0.91 in Burundi (where only 8 stations were analysed). The equation for

Java, with a wet tropical climate, gave much higher estimates for a given channel width than those for the other countries or the UK. This suggests that channel width is not sufficient to estimate the mean annual flood. It is also necessary to account for variations in geography, for example, as reflected in the climate.

Recent work on flood frequency estimation from channel geometry appears to have been concentrated in the USA. For example, Lawlor (2004) described sets of regression equations based on either BCW or active channel width, estimating design flows for return periods between 2 and 500 years in western Montana. The regression for the 2-year flow on BCW in the West Region of this area turned out to have an identical power term to that published in the FEH:

West Region of western Montana:  $QMED = 0.281 BCW^{1.98}$  (equation 3.2)

FEH Volume 3:  $QMED = 0.182 BCW^{1.98}$  (equation 3.3)

The return period for the bankfull discharge was found to range between 1.0 and 4.4 years, with a median value of 1.5 years. Lawlor (2004) also showed that both drainage area and mean annual precipitation were significantly related to bankfull discharge. This should serve as a reminder that the FEH regression equation for QMED based on catchment descriptors already contains terms that are strongly related to the bankfull discharge capacity. It is therefore possible that information on channel geometry, while significant in the absence of any catchment descriptors, could add little to the current equation's ability to predict QMED. This question has been investigated further within the FEH Local research (see Section 4.3).

More recently, a study by Modrick and Georgakakos (2014) used bankfull width alongside catchment area to improve estimates of flood flow for small mountainous catchments in southern California.

Studies by other researchers, such as Dodov and Foufoula-Georgiou (2004), Booker and Dunbar (2008) and Booker (2010), developed multiscale models to improve estimation of the coefficients describing power-law relationships between discharge and width. Incorporation of catchment area and climatic factors, such as the frequency of discharge, were found to improve model performance. However, the relationship between flow and channel dimensions has only been assessed in these studies to a Q10 percentile (Booker 2010), that is, well below a flood flow rate. Therefore, further assessment would be required to apply these relationships for derivation of flood flow estimates.

The limitations inherent in hydraulic geometry discharge relations have been highlighted by authors such as Soar and Thorne (2001), who looked at the simplification of a complex system in which the effects of various factors are difficult to separate.

Alternative approaches exist that use physically based equations (for example, friction factors) and channel hydraulics to determine river discharge – discussed by Kaplan-Henry (2007) and Soto and Madris-Aris (1994). Such an approach can overcome some of the issues inherent with existing methods that use regime theory, such as application to geomorphologically active reaches. However, it may not be straightforward to assign a return period to the discharge estimated using hydraulic methods. Such methods also require considerably more data than empirical approaches.

## 3.15 Examples of complex flood behaviour

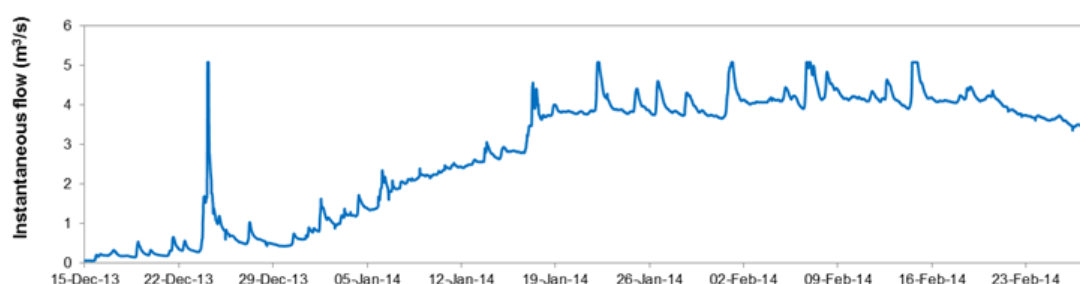
### 3.15.1 Ems at Westbourne, West Sussex

The difficulty of summarising the character and rarity of flooding on the highly permeable Ems at Westbourne catchment is captured by the hydrograph in Figure 3.4.

It is known that heavy rainfall alone led to the large flood on 24 December 2013. Later in the winter, the pattern of flooding was more characteristic of this highly permeable catchment, with high baseflow sustained over many weeks in January to February 2014, heightened at intervals by the flood response to individual storms. There is some evidence of bypassing in the array of 5 floods, each peaking (or truncated) at a flow of just over  $5\text{m}^3/\text{s}$ .

The challenge of flood estimation on this highly permeable catchment is highlighted by the following 3 features.

- The annual maximum flow is not well defined; most likely it occurred on 14 February 2014.
- The annual maximum flow does not do justice to the complex character of flooding in winter 2013 to 2014.
- The flood impact would have been very much greater had the December storm occurred a month or two later.



**Figure 3.4** Winter 2013 to 2014 flows in the Ems at Westbourne, West Sussex

Source: Solent and South Downs Area issue of JBA Consulting (2014)

A subtle approach is required to judge the (small) probability of an extreme flood arising from a severe storm occurring when the groundwater levels are unusually high, or an exceptionally severe storm occurring at other times. Continuous simulation modelling using stochastically generated daily rainfall might be a possible approach to representing this joint probability problem.

### 3.15.2 Letcombe Brook at Grove, Oxfordshire

Reed (2008) studied the flood risk from the Letcombe Brook at Letcombe Bassett, Wantage and Grove in Oxfordshire. One aspect of the study is relevant here. The flood risk from Letcombe Brook on its entry to Wantage is essentially that posed by a highly permeable catchment. Downstream, in the adjoining conurbation of Grove, the flood risk is more complex. Owing to contrasting land uses (rural and urban), slopes and soil types (highly permeable and moderately impermeable), Letcombe Brook is sensitive to an especially wide range of conditions, including the scenario of a severe convective storm on Wantage–Grove.

### 3.15.3 Ouse at York

Reed (2003) examined the meteorological signature in time and space of the 19 largest floods at York in the period 1881 to 2000. With only one exception (November 1951 event), the relative sizes of the floods at York were found to be consistent with the meteorological conditions experienced, once due account was taken of the many factors.

The most remarkable feature noted was the absence of any kind of signature flood event. Each of the 19 floods had its own unique character in terms of factors such as the spatial disposition of rainfall, the extent (or not) of long-term/medium-term/short-term antecedent wetness and the degree (if any) of snowmelt contribution. The complex behaviour reflects the unique character and configuration of the Ouse at York catchment.

The exceptionally long documented history of flooding at York was explored by Macdonald and Black (2010), who constructed an authoritative historical flood series spanning 1800 to 2000 and a more tentative one spanning 1263 to 2000. They found that unthinking fitting of a frequency curve to the gauged flood record would lead to a flood frequency curve unbounded above and to very considerable overestimation of the 100-year flood.

An explanation for the absence of outstandingly large floods in the historical and gauged record can be found in the strong attenuating influence of the very extensive floodplains of the Swale, Ure and Nidd tributaries close to their confluences upstream of York.

# 4 Development of procedures for enhanced use of local data

## 4.1 Summary of local data types

The scope of the project did not allow for a full investigation of all local data types and so development of new or improved procedures focused on the following data types:

- historical floods
- palaeofloods
- channel dimensions

The project also included some limited research on the use of additional types of hydrometric data such as baseflow index derived from daily mean flows (see Section 3.6.2).

The practitioner guide includes many pointers to good practice on the use of other types of local data, based on both published research and examples from consultancy work. The accompanying case studies include examples of using historical data, information on the impacts of recent floods, river levels, channel widths and information on catchment properties.

Table 4.1 summarises all the types of local data considered in the research and provides a commentary on the extent of their use in current UK practice (in the applied realm rather than academia), the potential for future increased uptake and the way in which they have been considered within the FEH Local research.

**Table 4.1 Summary of local data types considered in this project**

| <b>Data type</b>   | <b>Current UK practice</b>  | <b>Potential for future increased uptake</b>                 | <b>Coverage in FEH Local</b>   |
|--|---|--|--|
| River flow (gauges with uncertain high flow measurements or use of BFI data) | Widespread as good practice but not universal: many studies ignore any flow data from gauges not classed as suitable at least for QMED. | Moderate   | Review of current practice; included in user guidance but little new research. |
| River level  | Widespread availability and used to some extent   | Considerable   | Included in user guidance and case studies.                                    |
| Temporary gauges   | Rarely used   | Considerable: costly but need to weigh against the benefits. | Included for statistical framework for uncertainty estimation.                 |

| <b>Data type</b>   | <b>Current UK practice</b>  | <b>Potential for future increased uptake</b>  | <b>Coverage in FEH Local</b>   |
|--|---|---|--|
|  |   |   | Included in user guidance.   |
| Flood seasonality  | Widespread availability but rarely used.  | Moderate  | Review of current practice; included in user guidance and case study.  |
| Hydrograph widths  | Occasionally used   | Considerable for some types of study  | Review of current practice; included in user guidance.   |
| Climate variables beyond SAAR and FEH rainfall frequency | Hardly ever used  | Uncertain, perhaps minor  | No further work beyond review in previous chapter  |
| Groundwater levels                                       | Rarely used   | Moderate in limited circumstances   | Review of current practice; included in user guidance.   |
| Recent floods  | Often used for comparison with results of flood mapping studies.  | Need for guidance about how best to use.  | Review of current practice; included in user guidance and case studies.  |
| Historical floods  | Guidance readily available. Included on training courses. Information readily available for many rivers. Often mentioned in reports but only occasionally used to influence flood frequency estimation. | Considerable, although research and development of guidance needs to be accompanied by insistence from clients that methods are used. Development of software would help. | Investigation of statistical procedures (see Section 0 and Appendix B)<br><br>Proposal for improved access to data (Section 4.4)<br><br>Included in case studies and user guidance.          |
| Palaeofloods   | Information not easily available<br><br>Hardly ever used outside academia   | Considerable for studies of extreme floods  | In-depth review of current practice and research (Appendix A)<br><br>Investigation of statistical procedures (Section 0 and Appendix B)<br><br>Included in user guidance and 2 case studies. |
| Catchment properties beyond FEH descriptors              | Information readily available from maps and site visits but only occasionally used.   | Considerable for small catchments   | Soils, vegetation and land use covered by small catchments research  |



| <b>Data type</b>   | <b>Current UK practice</b>   | <b>Potential for future increased uptake</b>                     | <b>Coverage in FEH Local</b>   |
|--------------------|--|--|--|
|                    |  |  | High-resolution descriptors (Chapter 5)<br><br>Included in user guidance and case studies. |
| Channel dimensions | Hardly ever used apart from in conjunction with hydraulic modelling. | Unclear but worth investigating given ready availability of data | New procedures (Section 4.3)<br><br>Included in user guidance and case studies.            |

## 4.2 Statistical procedures for incorporating historical or palaeoflood data

### 4.2.1 Introduction

As discussed in Section 3.11, much research has been carried out on the best methods to estimate a flood frequency curve when both systematic and historical data are available at a location of interest. Different guidelines and approaches are employed across different countries, but despite having been promoted since the 1970s, historical data are not yet routinely included in flood risk assessment in the UK.

Within the FEH Local project, the applicability of some methods for British catchments was investigated via a large Monte Carlo simulation study. The practical consequences of the results of this extensive study are described in this section, while a more complete discussion of the design and results of the Monte Carlo study is provided in Appendix B.

A number of case studies showcasing the usefulness of historical data in reducing the variability of the estimated flood frequency curves are included in the practitioner guidance.

### 4.2.2 Statistical model – maximum likelihood

Results from the Monte Carlo study highlight that, for series with properties similar to those of the British records, the biggest gains in terms of reducing the uncertainty for estimated design events can be obtained using a maximum likelihood (ML) approach in which both the systematic (that is, gauged) and the historical data are included in the analysis.

The study shows that the method of L-moments, which performs quite well for short records, gives more variable results when historical data are included in the analysis. The use of approaches based on maximum likelihood is recommended when historical data are available. The maximum likelihood framework is briefly introduced here and more information is given in Appendix B.

If only systematic data at a gauging station are available for the site of interest, the sample of the annual maxima (AMAX) of gauged flow measurements is indicated as  $\mathbf{x} = (x_1, \dots, x_n)$ . It is assumed that each AMAX is the realisation of the same underlying statistical distribution  $f(x, \boldsymbol{\theta})$ . In the UK, this distribution is typically assumed to be the GLO, but other distributions like the GEV distribution or the Pareto Type 3 (PE3) distribution might be used in some cases. Each possible distribution is generally indexed by a set of parameters  $\boldsymbol{\theta}$  that need to be estimated from the data. In the maximum likelihood framework, estimates of  $\boldsymbol{\theta}$  are found by maximising the likelihood function  $L(\mathbf{x}, \boldsymbol{\theta})$  defined as:

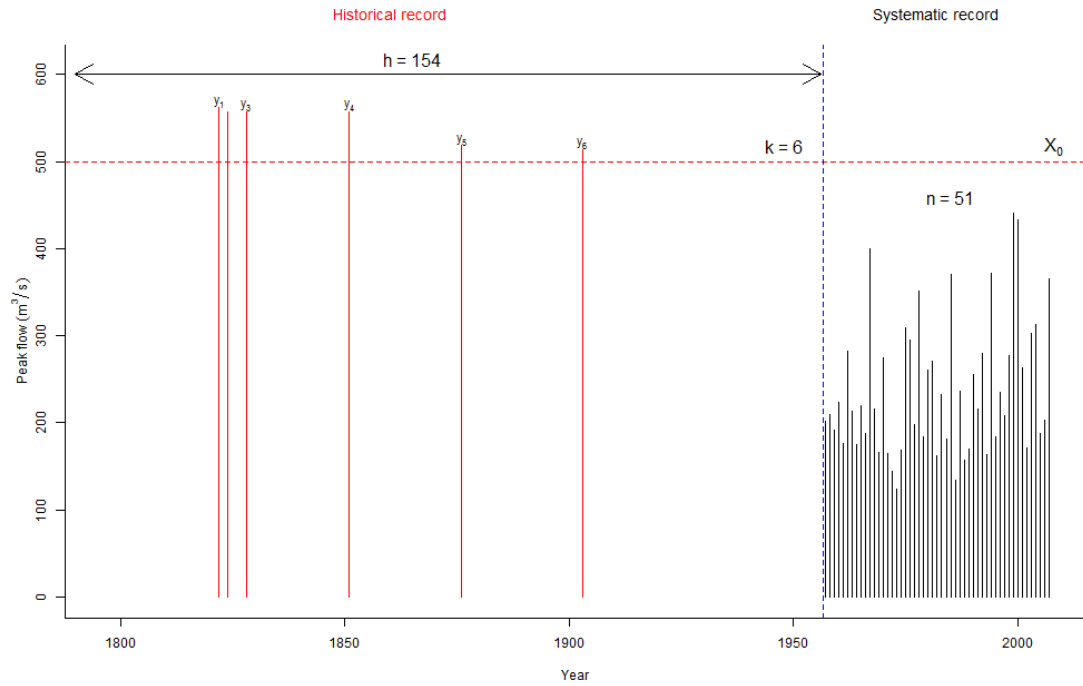
$$L(\boldsymbol{\theta}, \mathbf{x}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta})$$

(equation 4.1)

The maximisation is typically performed numerically, using general purpose optimisation algorithms.

For the framework to accommodate the inclusion of historical data, the following pieces of information are required (Figure 4.1).

- Gauged annual maximum flows  $(x_1, \dots, x_n)$ , referred to also as systematic data.
- Flow rates for  $k$  historical flood events, denoted as  $(y_1, \dots, y_k)$ , or at least information that  $k$  events occurred with a flow exceeding a threshold.
- The length of the historical period represented by the events,  $h$ .
- The threshold flow  $X_0$ , which is often named the perception threshold, since it corresponds to the threshold above which a flood would have been large enough to be noted in historical sources or leave recognisable signs across the catchment.



**Figure 4.1 Historical data example (River Wear at Durham)**

**Notes:** The graph shows a total of  $k = 6$  historical events (red bars) above the perception threshold  $X_0$  (dashed red line), recorded across the  $h = 154$  year-long historical period. The  $n = 51$  years long systematic record of gauged peak flows is also shown (black bars).

Assuming that all peak flow data follow the same distribution  $f(x, \theta)$ , in any given year the probability of exceeding the perception threshold  $X_0$  is  $(1 - F(X_0))$ , where  $F(x)$  indicates the cumulative distribution function. The number of events above the threshold in the historical period can then be represented as a binomial distribution. Starting from this consideration and other basic reworking of the distribution of the historical peak flow value, the likelihood to be maximised when historical data are available can then be written as:

$$L(\theta, \mathbf{x}, \mathbf{y}) = \prod_{i=1}^n f(x_i, \theta) * \prod_{j=1}^k f(y_j, \theta) * \binom{h}{k} F(X_0, \theta)^{h-k}$$

(equation 4.2)

It is often the case that no reliable information on the flow rates of the historical event can be obtained, but there is enough evidence to be confident that  $k$  historical flood events have exceeded a threshold flow  $X_0$ . In this case, the likelihood to be maximised becomes:

$$L(\theta, \mathbf{x}, \mathbf{y}) = \prod_{i=1}^n f(x_i, \theta) * \binom{h}{k} F(X_0, \theta)^{h-k} (1 - F(X_0, \theta))^k$$

(equation 4.3)

In all cases, maximisation of the likelihood is done using numerical optimisation routines. It is known that these routines might fail in some instances, or might be very sensitive to starting values, possibly owing to influential points on the sample under study or to some mis-specification of the model (for example, the wrong distribution

function is used to describe the data). Some trial and error might be needed to identify stable maximum likelihood results. It is good practice to verify that the assumed distribution is indeed suitable for the data under study, at least via graphical means such as Q-Q (quantile–quantile) plots and flood frequency curves.

### 4.2.3 Findings

As discussed in Appendix B, the simulation study showed that the inclusion of historical data can lead to a large reduction in the uncertainty around the estimated design event. This is particularly the case for very skewed data series for which information on a large number of historical peaks across a long historical period is available. The results also show that the gains are still present even when the parameters and the data used in the model are mildly mis-specified or incorrect. This does not mean that the efforts to correctly identify and quantify all necessary quantities needed in the estimation can be reduced, but it gives some reassurance that the use of historical data would still be beneficial even if some details are somewhat uncertain.

Based on the results of the Monte Carlo study, the following recommendations and guidelines are provided.

- The largest gains in terms of uncertainty reduction for estimates of the  $Q_T$  event are obtained when the perception threshold above which historical data is recorded corresponds to the  $Q_T$  value (see Section B.5). This could direct the effort in retrieving historical data when specific design events need to be estimated.
- The characteristics of the data available have some foreseeable effects on the estimates obtained when combining the historical record with a systematic record (see Section B.6). If the perception threshold  $X_0$  is not exceeded in the systematic record, it is very likely that the frequency curve obtained when including the historical data in the estimation procedure will be steeper than the one fitted to the systematic data only – more so for lower perception threshold and larger sample sizes.
- If the perception threshold has been exceeded for a large number of events in the historical record, the gains connected to the inclusion of the historical data would be larger when the information on the magnitude of the peak flow could be acquired (see Section B.7.1). If only a few (say, less than 5 for many practical cases) historical events are available, the precise number of threshold exceedances is the most important information to extract from the historical record.
- It is recommended that every effort is taken to ensure that complete information on the historical events is retrieved, with a correct understanding of the time covered by the historical information and as much information as possible on the number and size of the peak flows. Nevertheless, the estimation procedure tends to be fairly robust when a small proportion of the historical events are not present in the historical record (see Section B.7.2). Even if there might be some uncertainty around the correct number of threshold exceedances, if there is enough confidence that most of the historical events have been identified, the use of the historical records is recommended.
- When retrieving historical information, the characterisation of the actual time period covered by the historical information,  $h$ , should be carried out with as much care as possible. If no information is available to reliably quantify  $h$ , a statistical estimate can be taken to be twice the average time

interval between the historical events and the beginning of the systematic record (see Section B.7.3). If this estimate is smaller than the time between the first historical record and the beginning of the systematic record, this distance should be taken as an estimate for  $h$ . However, this should lead to questioning whether the dataset retrieved is complete and whether all data points included in the analysis are representative of the current flood risk.

- The value of the perception threshold  $X_0$  used in the estimation procedure can have a large impact on the final estimates. Its correct characterisation is a key step when integrating systematic data with historical records (see Section B.7.4).
- The use of historical data is discouraged in cases where a short systematic record (less than 10 years) is available and information on a very large historical flood is available. The resulting estimates are likely to be very biased (see Section B.4).

Based on these findings, recommendations have been included in the practitioner guidelines.

## 4.3 Estimating QMED from channel dimensions

### 4.3.1 Introduction

For a review of previous research into this topic, see Section 3.14. The FEH Local project included a small amount of research that aimed to build on previous work, updating it and examining the potential for estimating QMED using a combination of channel dimensions and catchment descriptors. There were 2 main tasks:

1. Re-do the analysis of Wharton et al. (1989) with QMED as the comparator, estimated both with and without the application of donor catchments. If estimates of QMED from channel capacity prove to be more accurate on a significant number of catchments, investigate these catchments to see whether they have any unusual features in common.
2. Expand the analysis of Wharton et al. (1989), in particular adding more small catchments. Investigate whether adding a measure of channel geometry to the QMED regression equation significantly increases the explanatory power of the regression.

The datasets and results are described below, after a discussion of the applicability of channel width methods.

### 4.3.2 Where to apply channel width methods

According to Wharton (1992), the channel geometry method for estimating stream flow is most suited to perennial streams with stable banks that are not easily widened by floods, such as upland streams with coarse armour or lowland streams with well-vegetated and cohesive banks. It is less likely to be accurate on flashy or ephemeral streams.

The regression on bankfull width provided in the FEH is not recommended for estimating QMED on:

- artificial channels

- strongly channelised rivers (unless the channel system has adjusted to the new flow regime)
- reaches with bedrock banks
- braided and geomorphologically active reaches
- reaches with large pools or locally steep gradients

It is also not recommended for use on streams where the channel width at bankfull is much less than 5m, as the data collected by Wharton et al. (1989) include only 3 channels narrower than 5m. This restriction probably excludes the majority of small catchments, at least in lowland areas.

Reaches selected for measurement of dimensions should be relatively straight or on stabilised reaches of meandering channels and at least 4–5 channel widths in length (Wharton 1992). Guidance on selecting cross-sections and measuring width is given by Wharton (1992).

### 4.3.3 Datasets

Various datasets were used or considered to carry out this analysis.

#### *Geraldene Wharton (GW)*

There are 75 sites, each with an associated bankfull channel width (BCW) and estimate of channel capacity, among other parameters. These are the 'bankfull' sites described by Wharton (1992), where dimensions were obtained from a new survey using consistent methods.

The survey work was carried out in 1987 to 1988. For a typical flood peak series 40–50 years long by the year 2015, 1990 is close to the mid-point of the record. The survey can therefore be taken as representative of channel conditions during the period of record, though it is acknowledged that channels will have changed at some locations.

There is an argument that channel dimensions should be related to flows recorded only in the period before the date they were surveyed, since the channel size reflects the flows that have shaped it. However, re-survey of channel dimensions was outside the scope of the project. It was decided that, on balance, the reduction in uncertainty in estimation of QMED thanks to the addition of 25 years of data would outweigh the disadvantages of the temporal mismatch. A discussion of trends in channel capacity and implications for flood risk can be found in Slater (2016).

A small number of slightly or moderately urbanised catchments were retained in the dataset for the analysis described below, under the assumption that the channel size had adapted to any increase in QMED resulting from urban development. None of the catchments is heavily urbanised.

The locations of the sites can be seen in Figure 4.4 later in this chapter. Of note is the almost complete lack of stations in central England and East Anglia.

### *River Habitat Survey (RHS)*

This dataset contains channel survey data for thousands of sites around the country and has a wide range of information for each site, including bank channel width and information on whether the site exhibited obvious unnatural channel modification. Only RHS sites located on the same stretch of river as a NRFA gauging station, with no intervening tributaries, were considered.

### *National River Flow Archive (NRFA)*

This provided annual maximum flow data from which an up-to-date estimate of QMED was calculated for 73 of the 75 sites in the GW dataset.

### *Ordnance Survey Open Rivers shape file*

This Ordnance Survey (OS) product was used to provide a river ID for each NRFA gauging station and RHS site, thus allowing them to be joined using GIS software. This allowed those RHS sites not close enough to gauging stations to be discounted from the analysis.

After some trials, a maximum distance of 1km was imposed. Where there were multiple RHS sites within this distance, their channel widths were averaged to give a measurement more representative of the reach.

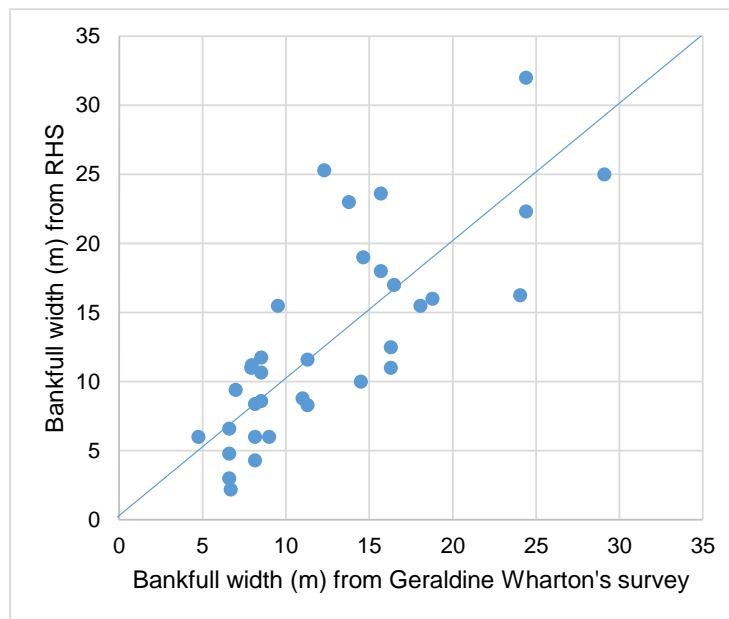
### *LiDAR*

Light detection and ranging (LiDAR) data offer a possible route for rapidly obtaining BCW measurements at many more gauging stations than the 75 available from the GW dataset. However, it does not seem likely that remotely sensed data will be able to satisfy all the criteria for selection of cross-sections and field survey techniques defined by Wharton (1992). Particularly on narrow watercourses, 2m resolution LiDAR will not be sufficiently detailed to provide information on the BCW. It may be more promising for wide channels, although it would need to be carefully combined with information from other sources such as the RHS dataset and flood defence databases – National Flood and Coastal Defence Database (NFCDD) or the Environment Agency's Asset Information Management System (AIMS) – to ensure that only natural reaches are selected.

## **4.3.4 Comparing GW and RHS datasets**

Although it was suggested that the RHS dataset could provide a potential way to add many more sites to the analysis, the RHS was not designed as a topographic survey. Channel dimensions were merely estimated and the field work was designed to be conducted by trained non-specialists rather than geomorphologists.

With these concerns in mind, the RHS data were compared with the BCW from the GW dataset (Figure 4.2). As expected, given the differences in methodology and timing of the surveys, there was a difference between the 2 datasets. Differences of up to 10m in width can be seen at some sites in the scatter plot, although at most sites the measurements are within 5m.



**Figure 4.2 Comparison of bankfull channel width from the RHS and GW datasets**

The mean absolute difference is 3.6m, which is 28% of the mean channel width. This discrepancy is unacceptably large. The GW data were carefully surveyed using consistent methods designed for the express purpose of measuring hydraulic channel characteristics and therefore the remainder of this document focuses solely on the GW dataset.

Only 3 sites in the GW dataset have a channel narrower than 5m. The narrowest width is 3.6m. This restricts the utility of the results on narrow channels, such as might be expected on lowland streams in small catchments. However, there is a reasonable number of small catchments in the dataset, with 13 catchments under 25km<sup>2</sup>.

Since the accuracy of the RHS data is thought to be better for narrower rivers, the addition of some channels narrower than 5m from the RHS dataset was considered. Only 7 NRFA gauging stations were found with RHS data sufficiently close and channel widths under 5m. After examining these, it was decided not to include them in the analysis. In most cases, widths were rounded to the nearest 1m, which does not inspire confidence in the accuracy of the measurements.

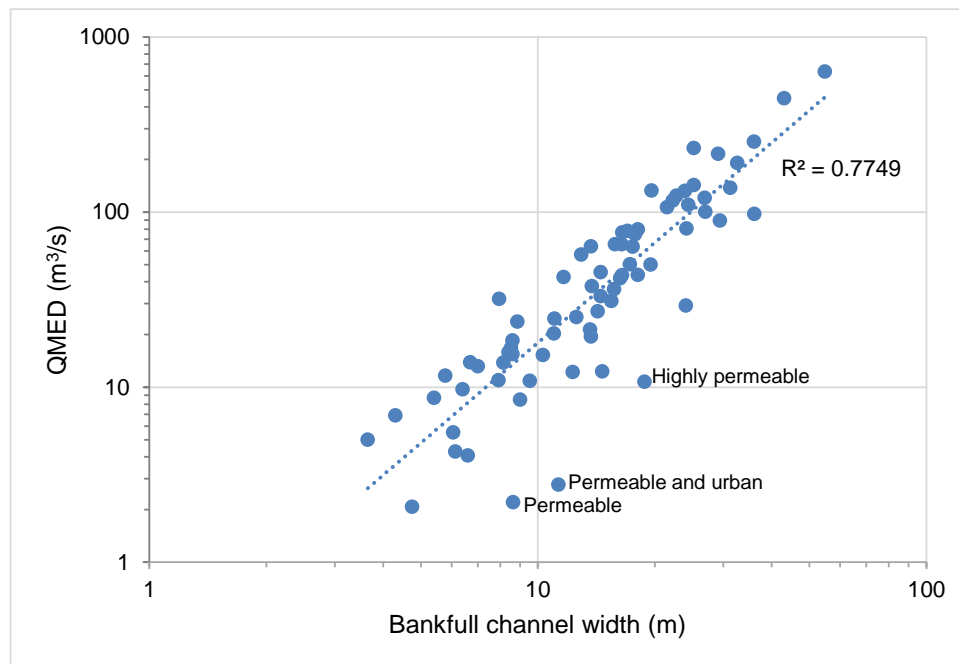
There is also a shortage of very wide channels in the GW dataset, with only 2 wider than 40m. This is less of a concern because wide rivers are more likely to have flow gauges, which can be expected to act as useful donor sites, reducing the uncertainty in estimation of design flows from catchment descriptors.

#### 4.3.5 Regression for QMED on channel dimensions

The regression equation given in Section 5.2 of the FEH was updated using QMED estimates made from the latest NRFA peak flow dataset (version 3.3.4), which has 18 more years of data than that available for the FEH research. The analysis included 73 sites from the GW dataset, a few more than the 65 used to develop the FEH equation.

The fitted regression line is shown in Figure 4.3 and compared with the FEH equation in Table 4.2.





**Figure 4.3** Revised regression of QMED on channel width

**Table 4.2** Comparison of regression equations based on channel width

| Version                      | Equation                   | $R^2$      | Factorial standard error ( <i>fse</i> ) |
|------------------------------|----------------------------|------------|---|
| FEH (Equation 5.1, Volume 3) | $QMED = 0.182 BCW^{1.98}$  | 'Over 80%' | 1.73                                    |
| New                          | $QMED = 0.226 BCW^{1.899}$ | 77%        | 1.60                                    |

As shown in Table 4.2, the coefficients of the 2 equations are similar;  $R^2$  has been reduced, but so has *fse*. However, *fse* remains considerably higher than that for the current FEH regression equation for predicting QMED from catchment descriptors, for which *fse* is 1.43 (and  $R^2$  is 0.94).

It appears that the largest outliers are on highly permeable catchments. There may be a tendency for chalk or limestone streams to have different geometric characteristics, perhaps being shallower and wider, or to have a channel-forming flood flow higher than QMED. According to Harvey (1969), the recurrence interval of the bankfull discharge on a baseflow-dominated stream, the Wallop Brook in Hampshire, was considerably higher than 2 years, being between 5 and 10 years at most sites. One explanation put forward was that, on a baseflow stream, the annual flood may not be competent to cause scour of the banks. It may be that including a measure of catchment permeability will improve the regression; this is explored further in the following section.

Other options, such as regression on other measures of channel geometry, were not explored. Bankfull width has the advantage of being easily measured, and performed well compared with other variables in the analysis reported by Wharton (1992).

Similarly, no attempt was made to predict flood quantiles for longer return periods, as there is less of an expectation that there is a geomorphological relationship between channel capacity and extreme flood magnitudes. In addition, estimates made from at-site analysis of flood peak data become more uncertain for longer return periods.

### 4.3.6 Comparing best QMED predictor

Questions that may be of interest to practitioners include:

1. Can channel dimensions give an improved estimate of QMED compared with catchment descriptors?
2. If so, on what types of catchment?
3. Is the improvement maintained even when the catchment descriptor estimate of QMED is adjusted using a donor site?
4. How could information from catchment descriptors and channel dimensions be combined to give a reduction in the uncertainty of estimating QMED?

The 4 questions are discussed in turn below. Evidence to help provide the answers was compiled by analysis of the 73 sites discussed above. At each site, QMED was estimated using:

- the current FEH regression equation on catchment descriptors (from Environment Agency 2008), including an urban adjustment where necessary (**CDs**)
- the above plus a donor adjustment (**Donor**), selecting the next nearest gauged catchment as a donor, treating the subject catchment as ungauged
- regression on the BCW, using Equation 5.1 from FEH Volume 3 (**BCW:FEH**)
- the new version of the channel width regression, derived as explained above (**BCW:new**)
- a new regression equation combining channel width and catchment descriptors, described below under Question 4 (**BCW + CDs**)

#### *Question 1: Can channel dimensions give an improved estimate of QMED?*

The predictions were compared with the best estimates of QMED made directly from the annual maximum flow data at each gauged site. The results are summarised in Table 4.3. The best predictor is defined as the method that gives the smallest absolute value of:

$$\log (\text{QMED}_{\text{predicted}}/\text{QMED}_{\text{obs}}).$$

**Table 4.3 Comparison of QMED prediction from catchment descriptors and from channel width**

| Estimation method for QMED   | CDs  | BCW:FEH | BCW:new |
|--|------|---------|---------|
| Number of sites where method was best predictor                                    | 50   | 13      | 10      |
| Root mean square error of predictions  | 35.8 | 37.8    | 41.9    |
| <b>Distribution of residuals, expressed as ratio of predicted to observed QMED</b> |      |         |         |
| Largest overprediction   | 6.27 | 7.95    | 8.11    |
| Upper 95th percentile  | 1.69 | 3.14    | 3.09    |
| Median   | 1.03 | 0.87    | 0.87    |
| Geometric mean (the bias of the method)  | 1.00 | 1.00    | 1.00    |
| Lower 95th percentile  | 0.59 | 0.49    | 0.51    |
| Largest underprediction  | 0.22 | 0.35    | 0.36    |

CDs were found to give the best estimate of QMED at 50 out of the 73 sites (68%); this method shows the lowest root mean square error (RMSE) overall. This leaves 32% of the sites where one or other (usually both) of the BCW methods gave the best estimate. For this dataset, the FEH regression on BCW performs rather better than the new version of the regression and indeed its RMSE is only a little above that for the CDs regression. None of the methods show any bias overall. This is to be expected for the BCW approaches, which have been developed and assessed using an identical dataset.

Looking at the largest errors, the largest underprediction is from CDs and the largest overprediction from the new BCW regression.

Despite being based on only one variable, compared with the 4 catchment descriptors used in the QMED regression (along with urban extent where necessary), the BCW method gives a better estimate of QMED at around a third of the sites. It is worthy of further investigation.

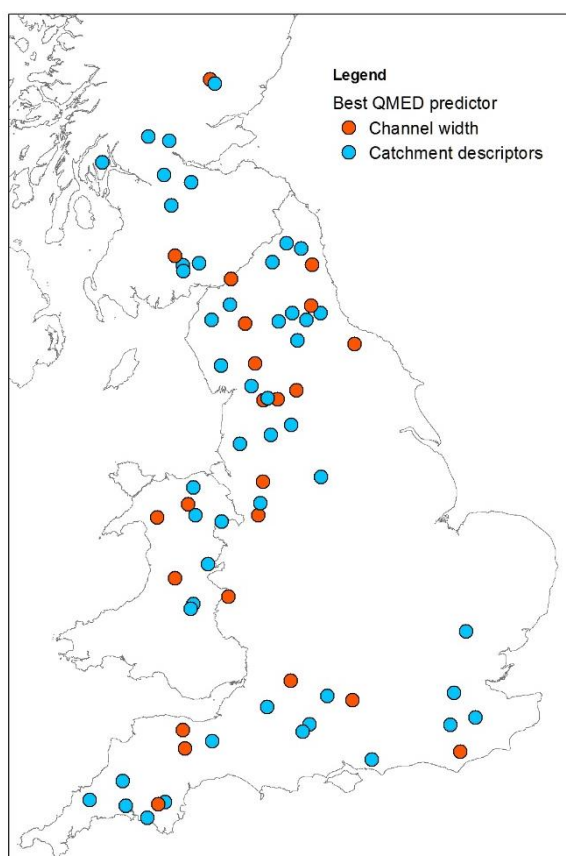
## Question 2: Types of catchment where the channel dimension method performs well

If it was possible to identify some types of catchment for which BCW consistently gave better estimates of QMED than CDs, practitioners could be recommended to give some weight to the results of the BCW method on such catchments.

The 73 catchments were divided into 2 groups: the 50 where CDs perform best and the 23 where they do not. Within each group, the FEH catchment descriptors were averaged – both those used in the estimation of QMED and those that are not.

No significant differences were observed in the physical properties of the 2 groups. They were similar in terms of size, length, steepness, altitude, soil properties, rainfall, soil wetness and extent of floodplains. Neither was there any clear difference in their location, as can be seen in Figure 4.4.

It was concluded that there is no straightforward way of formulating a recommendation on when to prefer the result obtained using the BCW method.



**Figure 4.4** Spatial distribution of gauging stations in the BCW dataset, showing which method gave the best prediction of QMED

### Question 3: Effect of introducing a donor adjustment

When the donor catchment adjustment is introduced, the results change (Table 4.4).

**Table 4.4 Comparison of QMED prediction from catchment descriptors and from channel width**

|  | <b>CDs with donor</b> | <b>BCW:FEH</b> | <b>BCW:new</b> |
|--|-----------------------|----------------|----------------|
| Number of sites where method was best predictor                                    | 54                    | 10             | 9              |
| Root mean square error of predictions  | 25.4                  | 37.8           | 41.9           |
| <b>Distribution of residuals, expressed as ratio of predicted to observed QMED</b> |                       |                |                |
| Largest overprediction   | 6.74                  | 7.95           | 8.11           |
| Upper 95th percentile  | 1.64                  | 3.14           | 3.09           |
| Median   | 1.05                  | 0.87           | 0.87           |
| Geometric mean (the bias of the method)  | 1.04                  | 1.00           | 1.00           |
| Lower 95th percentile  | 0.66                  | 0.49           | 0.51           |
| Largest underprediction  | 0.23                  | 0.35           | 0.36           |

As expected, the performance of the CDs method improves when the donor adjustment is introduced. It now gives the best prediction of QMED at 74% of sites and the RMSE has dropped considerably compared with Table 4.3. However, it makes little difference to the worst underprediction from CDs.

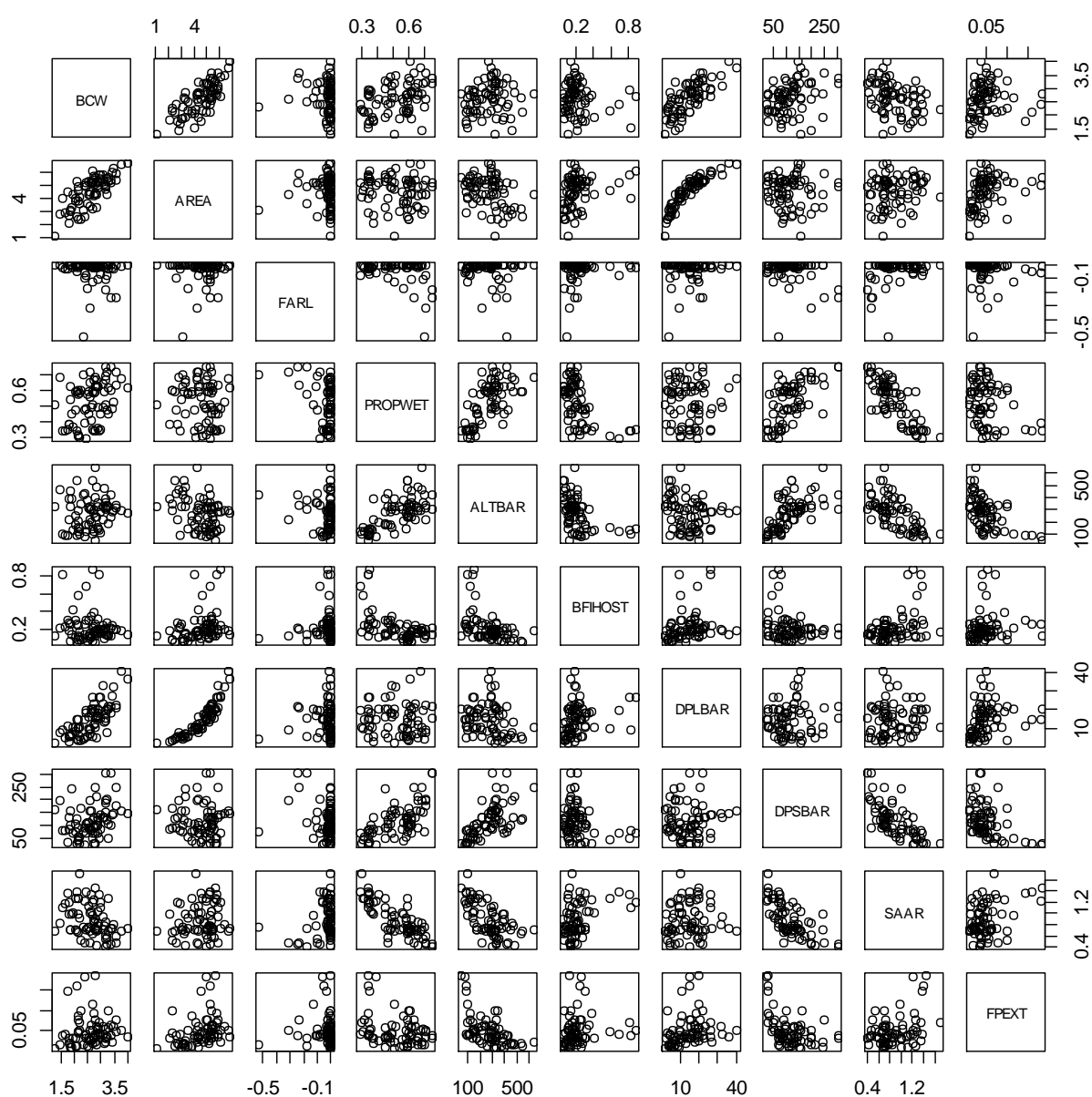
Introducing the donor catchment adjustment for QMED predicted from catchment descriptors might be viewed as an unfair comparison, since no equivalent correction is being made to QMED predicted from channel width. Although such an adjustment would be theoretically possible, there is not currently a national dataset of channel widths for every gauge location and so the information required to develop that type of adjustment procedure does not exist. This point should be borne in mind when interpreting the results: the performance of the channel dimensions method is painted in a poor light when compared with the donor adjustment method.

## Question 4: Potential for combining catchment descriptors and channel dimensions

### Exploratory analysis

Since it was not possible to identify any catchment types on which the BCW approach consistently performed better than CDs for estimation of QMED, a remaining question is whether the prediction of QMED could be improved by merging information from catchment and channel descriptors.

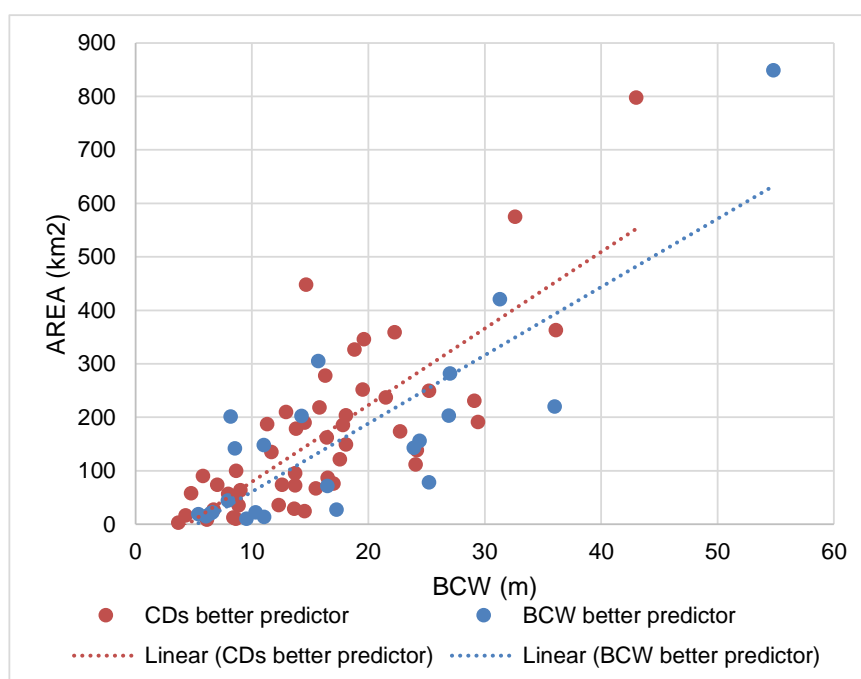
As a first stage, the correlation between BCW and (transformed) FEH catchment descriptors was investigated, as illustrated in the scatter plot matrix in Figure 4.5. There is a marked correlation between BCW and AREA, as might be expected (Kendall's correlation coefficient  $\tau = 0.55$ ). There is a similar correlation with mean drainage path length, DPLBAR.



**Figure 4.5 Scatter plot of FEH catchment descriptors and BCW**

No pattern was found to suggest that sites where BCW performs better than CDs are those where BCW is unusually wide or narrow compared with what might be expected from catchment descriptors such as AREA. As can be seen in Figure 4.6, there is little difference in the relationship between BCW and AREA for sites where BCW gives the best prediction of QMED and sites where CDs give the best prediction.

Neither was any relationship found between the performance of the BCW approach and the size of the resulting QMED estimate relative to the estimate from CDs.



**Figure 4.6 Relationship between BCW and catchment area**

These checks appear to rule out the possibility of developing simple guidelines such as 'Prefer the BCW approach on catchments where BCW is unusually large for the catchment size' or 'Prefer the BCW approach where it gives a larger estimate of QMED'.

A more sophisticated approach is required, such as one where BCW (or other channel dimensions) is included as a candidate variable in a regression of QMED on catchment descriptors. This is not currently possible using the full national peak flow dataset owing to the lack of readily available and reliable channel dimension data for the majority of gauging stations, but it could be considered for future work.

It is therefore unlikely that any regression that includes BCW will improve on the existing FEH catchment descriptor model. However, it can be expected that a regression including BCW and catchment descriptors could improve on the simple regression using only BCW (as described in Section 4.3.5) and provide an alternative approach to estimating QMED.

### Developing a new regression

Catchment descriptors were mathematically transformed in the same way as in the regression equation given in Environment Agency (2008), that is, log transformations for QMED, AREA and FARL, inversion for SAAR (1000/SAAR) and raising BFIHOST to the power of 2. BCW was log-transformed. A step-wise multiple linear regression was carried out (that is, variables were introduced one at a time).

The performance of the regression equations was measured by the proportion of variance explained ( $R^2$ ), the Akaike Information Criterion (AIC, which gives an indication of when there are redundant variables included in the equation) and the significance of the  $t$ -test (which checks whether each coefficient is significantly different from zero).

There is a marked tendency for the BCW regression to overestimate QMED on highly permeable catchments (see Figure 4.3). When adding one FEH catchment descriptor to the regression, the largest improvement can be achieved by adding BFIHOST, for which  $R^2$  increases from 0.77 to 0.84.

The best 3-variable equation uses BCW, BFIHOST and AREA, and gives an  $R^2$  of 0.86. Given the correlation between BCW and AREA, it is worth considering the inclusion of a term in the regression that represents the interaction between these 2 variables. This was tried but it did not improve the regression. There is some concern about the inclusion of AREA given its correlation with BCW.

The best 4-variable equation uses BCW, BFIHOST, AREA and SAAR, with an  $R^2$  of 0.89.

With all 5 variables (4 from the FEH regression plus BCW),  $R^2$  is 0.91. All 5 variables are significant at the 99% confidence level, and the AIC is lower than for the regressions with fewer variables. The  $fse$  value is 1.385. All descriptors have coefficient signs (+ or -) that would be expected from knowledge of their physical influence on QMED.

$R^2$  is not as high as the 0.95 achieved in the FEH regression (Environment Agency 2008), using a much larger dataset. But the  $fse$  is lower than the 1.431 associated with the FEH regression (that is, the estimate of QMED is more certain). However, this new regression is based on a much smaller dataset than that used for the FEH regression. For comparison, an FEH-style regression using the 4 catchment descriptors calibrated on this dataset achieves  $R^2 = 0.90$  and  $fse = 1.413$  (that is, a marginally lower performance than when BCW is included).

The BCW + CDs regression model is:

$$\ln QMED = 0.9271 + 0.661 \ln BCW - 2.814 BFIHOST^2 + 0.6028 \ln AREA + 2.181 \ln FARL - 1.324 (1000/SAAR)$$

that is:

$$QMED = 2.527 BCW^{0.661} 0.0600 BFIHOST.BFIHOST AREA^{0.6028} FARL^{2.181} 0.266^{(1000/SAAR)}$$

(equation 4.4)

### Performance of the new regression

Figure 4.7 compares the results from the new BCW + CDs regression with those from the BCW:new regression.

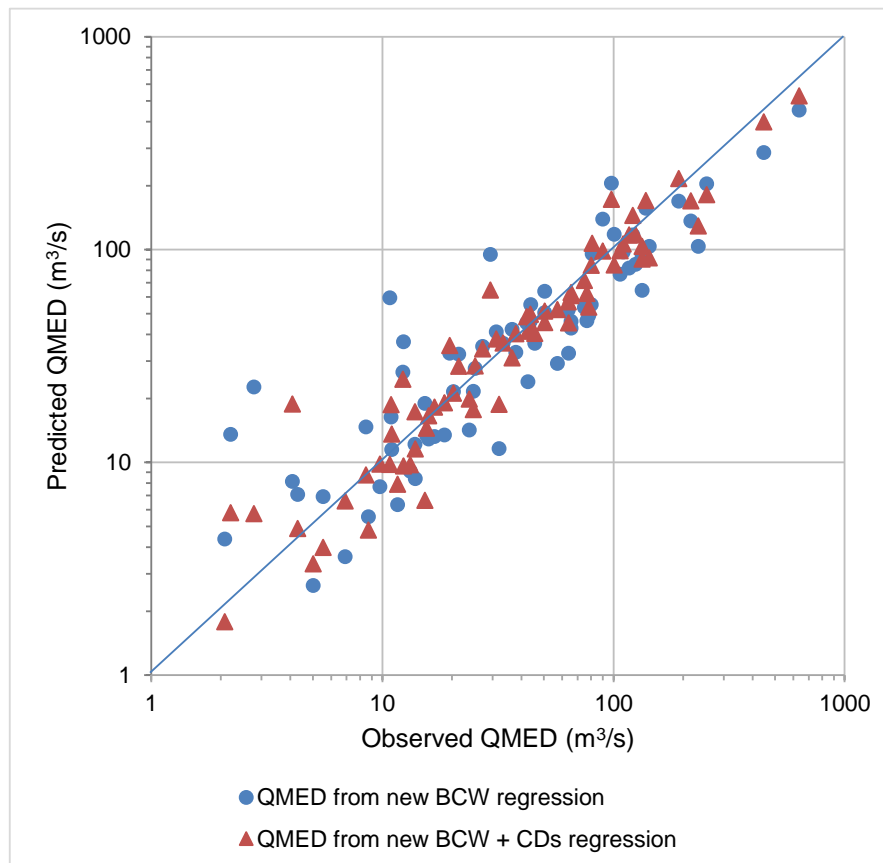
An urban adjustment was applied to the results from the BCW + CDs regression using the procedure from Kjeldsen (2010); this increased the estimates by 6–17% on 6 slightly or moderately urbanised catchments and by 40% on one slightly urbanised but highly permeable catchment. This urban adjustment was developed specifically for as-rural estimates of QMED produced by the FEH CDs regression and so its application to another regression is not necessarily valid, particularly in this case where the effects of



the upstream urbanisation will be already present in the channel width used to estimate QMED.

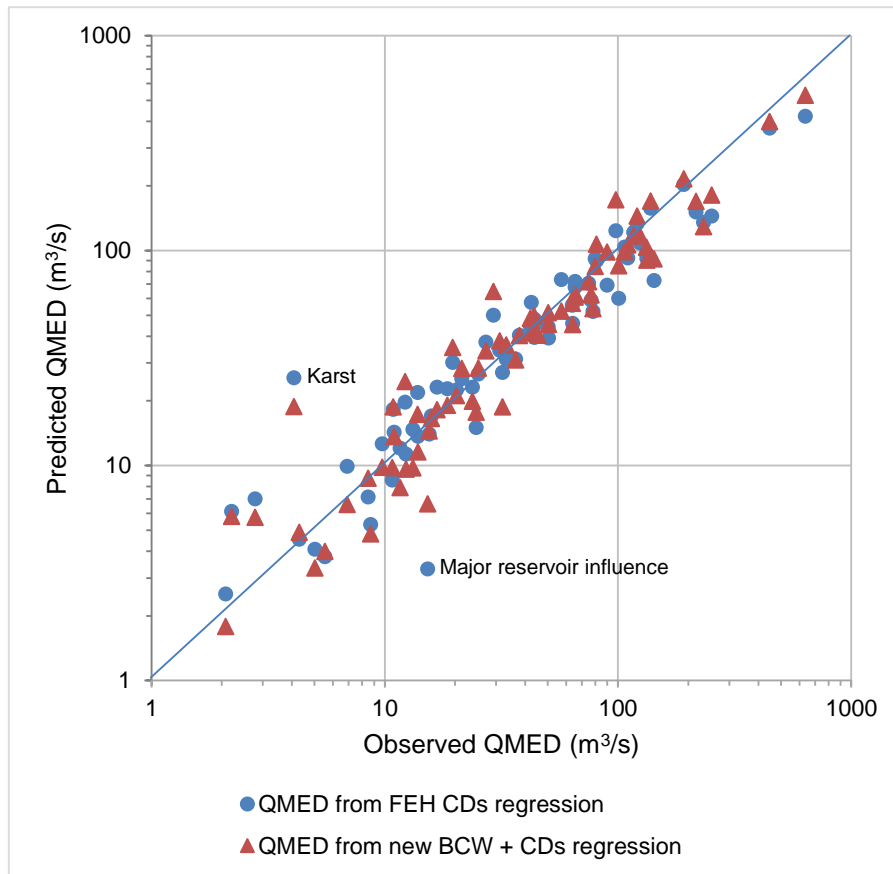
Despite these concerns, the urban adjustment does appear to improve the estimation of QMED on urbanised catchments and so has been retained for the purposes of this exploration. Although there may be an argument for excluding urban catchments from this analysis (as is typical in FEH research), there is a need to develop procedures that are applicable on urban catchments, since so many flood studies take place on such catchments. Some more comments on this are given in Section 4.3.7.

Figure 4.7 shows that there is a reduction in scatter when CDs are added to the BCW regression.



**Figure 4.7 Comparison of predicted and observed QMED from 2 regression models**

Figure 4.8 compares the FEH CDs regression and the BCW + CDs regression. In both cases, urban adjustments are applied. The 2 largest outliers in the FEH CDs regression appear to be improved in the BCW + CDs results.



**Figure 4.8 Comparison of predicted and observed QMED from 2 regression models**

As shown in Table 4.5, the RMSE of the results predicted by the BCW + CDs regression is considerably lower than that for the FEH CDs regression, thanks in part to the major improvements to the largest outliers. The BCW + CDs regression manages to reduce the RMSE to be similar to that achieved by CDs with donor adjustment. In terms of numbers of sites, the method that most often gives the best estimate of QMED is BCW + CDs, rather than CDs with donor adjustment.

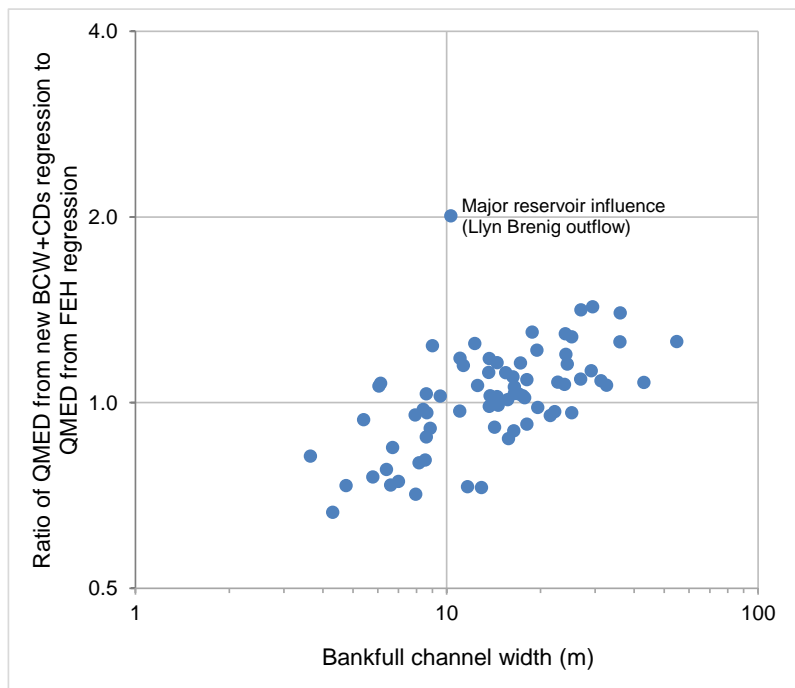
This result is potentially important, as it could imply that inclusion of channel geometry data has the potential to improve QMED prediction at least as much as carrying out data transfer from a donor catchment. In this set of catchments, it also reduces the largest over and underpredictions much more effectively than data transfer does.

**Table 4.5 Comparison of performance of 3 regression equations for QMED**

| Regression equation  | CDs                        | CDs with donor | BCW + CDs |
|--|----------------------------|----------------|-----------|
| Number of sites where method was best predictor                                    | Not included in comparison | 30             | 43        |
| Root mean square error of predictions  | 35.8                       | 25.4           | 26.1      |
| <b>Distribution of residuals, expressed as ratio of predicted to observed QMED</b> |                            |                |           |
| Largest overprediction   | 6.27                       | 6.74           | 4.61      |
| Upper 95th percentile  | 1.69                       | 1.64           | 2.08      |
| Median   | 1.03                       | 1.05           | 0.98      |
| Geometric mean (the bias of the method)  | 1.00                       | 1.04           | 1.02      |
| Lower 95th percentile  | 0.59                       | 0.66           | 0.63      |
| Largest underprediction  | 0.22                       | 0.23           | 0.43      |

Figure 4.9 investigates something likely to be of interest to practitioners: the change in QMED between the FEH CDs and BCW + CDs equations. It can be seen that, although there is no way of predicting the exact change solely on the basis of channel width (because the new regression is not just an adjustment to the FEH CDs procedure), there is a general tendency for QMED to decrease on narrow channels and to increase on wide channels. In this dataset, QMED nearly always decreases on channels narrower than 8m and nearly always increases on channels wider than 25m.

The obvious outlier is a site at the outlet of Llyn Brenig Reservoir, where FARL = 0.58. QMED estimated from BCW + CDs is double the estimate from CDs here, which brings it closer to the gauged value. In practice, for a site with this degree of reservoir influence, if ungauged, design flows should be estimated from rainfall-runoff methods with reservoir routing.



**Figure 4.9** Change in QMED between the FEH and the BCW + CDs regressions

#### Sensitivity to measurement errors

Even when the instructions for survey techniques in Section 4.3.1 are followed, measurement of BCW will not be exact. In situations where the location of the bankfull level is unclear, it is easy to envisage an error of, say, 1m in the measurement, even when averaged over several sections. A change of 1m in the value of BCW alters the estimated QMED by around 15% on average when it is estimated solely from BCW and by around 5% on average when estimated from BCW and CDs.

#### 4.3.7 Conclusions and recommendations on channel geometry

The results of the analysis described above indicate that channel geometry has the potential to improve the estimation of design flows.

If QMED is estimated solely from BCW, the result is, on average, not as accurate as that obtained from FEH catchment descriptors. At around a third of sites, however, the BCW method gave a better estimate of QMED. Unfortunately, no way has been found of detecting the type of situation where the BCW method is more reliable.

When BCW is combined with catchment descriptors in a new regression, the result appears to be more accurate (on average) than that obtained solely from catchment descriptors. While this might not be a surprising result, and the degree of improvement in the overall statistical performance of the regression is fairly slight, the new regression does appear to be able to achieve significant improvements in the estimation of QMED at a small number of outlier sites. This is important because these are the types of sites for which there is an expectation that local data could be particularly valuable. Furthermore, it appears that the new regression gives a better estimate of QMED more often than the use of catchment descriptors with a donor adjustment.

The analysis was based on a limited dataset of only 74 gauging stations. This is much smaller than the set of 602 catchments from which the current FEH regression equation for QMED was developed (Environment Agency 2008). It is recommended

that any future update to the FEH statistical method should include compilation of a reliable dataset of channel geometry data for all gauging stations, so that measures of channel geometry can be included as candidate variables in the regression or an alternative regression.

In the meantime, there is now a provisional alternative procedure available for estimation of QMED that combines catchment descriptors and channel width. It is not suggested that this new regression equation is applied routinely, since it has been developed from such a small dataset. However, in cases when there is doubt or concern over the accuracy of an estimate from FEH methods (for example, where other types of local data conflict with the results), one way forward would be to measure an average bankfull width and then apply the new regression to provide an alternative estimate of QMED. This would only be applicable on natural channels that satisfy the criteria listed in Section 4.3.1. Where there is little difference between the alternative estimates, this exercise should help in reinforcing confidence in the FEH result. If the new regression yields an increased estimate of QMED, this might be preferred for some types of project where a conservative answer is desirable.

The regression is not recommended for application in central England or East Anglia because of the absence of calibration sites in those areas (Figure 4.4).

Further analysis could be considered to:

- check the impact of excluding the urbanised catchments in the development of the regression
- investigate the utility of LiDAR data along with aerial photography for estimation of channel widths

## 4.4 Proposal for improving access to local data

Although the development of a new system to deliver local data to practitioners is outside the scope of the FEH Local project, it is envisaged that a future implementation would have greatest benefit if linked to the NRFA, which already holds the core peak flow data used for flood estimation in the UK. A feasibility study was therefore conducted on the development of a Local and Historical Flood Data Archive and its integration with the NRFA. Appendix C provides a report on the study and its main findings are summarised below.

A future implementation of the new system and its supporting procedures would need to have the functionality to collate, store, quality control and provide access to the data with a UK-wide remit. The host organisation and project team would need to consider options for the content and technical structure of the system and how it might be set up, operated and maintained. Business planning considerations would include the hosting, funding and governance of the system and the mitigation of risks associated with its use. This feasibility study outlines a number of options in each of these areas and makes recommendations as to the most beneficial.

With regard to archive content, there is a broad range of data types that could be of use to FEH practitioners in validating and refining their estimates of peak flows and flood frequency at their site of interest. These include:

- existing estimates of peak flows and related information
- hydrometric measurements additional to those already available on national archives

- catchment information such as local amendments to FEH catchment descriptors and changes in hydrological response over time

The feasibility study proposes that:

- hydrometric measurements are best hosted by existing national archives
- catchment information not pertaining to a specific event should be outside of the scope of the new system

The recommendation is therefore that the system contains estimates of flood events, quantitative records of flood extents and levels, and the raw observational information (such as photographs) from which they are derived.

With regard to the potential for integration of the new system with the NRFA, full integration into the NRFA is likely to be the most effective option, both with regard to costs and to engagement. This will exploit existing skills and technical infrastructure, and allow practitioners to discover and analyse the available information efficiently.

In setting up the system, funding would be required to establish the systems and procedures, and then to populate the empty archive with a core set of data. While it would be possible to postpone the population until after the launch, it is recommended that the system is established along with supporting policies and procedures and is populated with a core set of data by a consortium of organisations so as to maximise initial user engagement.

Post-launch, secure and sustained funding would be required for the ongoing operation and maintenance of the system. For data collation and quality control, it is recommended that users submit data, which are quality controlled by central or federated teams of experts, who may also be engaged in proactive collation of data.

For ongoing stewardship of the data, it is recommended that archive content is improved by the processing of user data quality flags and by rolling review. It might also be extended by an active programme of extraction of structured data from unstructured records.

Table 4.6 highlights the recommendations in 7 key areas that would require decisions in a future implementation of the system.

**Table 4.6 Recommendations for a new system to deliver local data to practitioners**

|  |   |
|--|---|
|  |   |
| <b>Archive content</b>   | The system contains estimates of flood events, quantitative records of flood extents and levels, and the raw observational information (such as photographs) from which they are derived.                   |
| <b>Back-end database</b>   | The system is held on local storage with in-house systems support, and contains both structured and unstructured data. Future phased expansions should accommodate growth in size.                          |
| <b>Front-end user interface</b>                                  | An advanced user interface allows federated searches across different (or integrated) databases, with options for data display and export, and tools for authorised users to add and amend records.         |
| <b>Integration with existing UK national flood data archives</b> | The remit of the NRFA is extended to host the new system, with the option of integrating all relevant records from the CBHE.  |
| <b>Setting up the system</b>                                     | The system is established along with supporting policies and procedures, and is populated with a core set of data by a consortium of organisations.   |
| <b>Data collation and quality control</b>                        | Users submit data, which are quality controlled by central or federated teams of archive staff, who may also be engaged in proactive collation of data.   |
| <b>Data stewardship</b>  | Archive content is improved by the processing of user data quality flags and by rolling review, and may also be extended by an active programme of extraction of structured data from unstructured records. |

In conclusion, the delivery of an archive to meet the aims of the FEH Local project is eminently feasible, given clearly defined limits on the data types to be included and secure funding, with a national remit for the establishment and long-term operation and maintenance of the system.

The position of the NRFA as the UK's central archive of peak flow data, with proven capability in database and web development and in hydrological data stewardship, and with established partnerships with the primary measuring authorities, commends it as a suitable host.

# 5 High-resolution catchment descriptors

## 5.1 Introduction

The FEH Local research included a feasibility study of the development of high-resolution FEH spatial datasets. It involved the following tasks:

- (a) Comparing methods for producing flow grids. This task considered the software options available for developing a new high-resolution drainage grid and associated catchment descriptors.
- (b) Creating a new high-resolution drainage grid from which existing FEH catchment descriptors can be derived.
- (c) Exploring new high-resolution catchment descriptors for use in the FEH. This was done in 2 stages:
  - Level 1: Using the new outflow grid, for greater positional accuracy, in combination with the existing source datasets
  - Level 2: Using the new outflow grid in combination with higher resolution or improved source datasets.
- (d) Producing an evaluation version of a  $Q_T$  point dataset (that is, automated calculations of design flows) for pilot catchments.

The principal aim of this study was to explore the potential for using new sources of topographic and other spatial datasets that were not available when the FEH was originally developed. It was envisaged that there could be potential benefits from defining catchment descriptors at a high resolution, particularly for small catchments where FEH catchments are currently subject to a lower size limit of  $0.5\text{km}^2$  and are thought to be more prone to errors in catchment boundaries or descriptors such as soil properties owing to the coarse resolution of the FEH spatial datasets.

## 5.2 Drainage grid

The aim of this task was to develop a process by which catchment boundaries can be calculated from a high-resolution terrain model for regularly spaced points along the watercourse network, down to a minimum catchment area of  $0.2\text{km}^2$ .

The minimum area in the current FEH dataset is  $0.5\text{km}^2$ . Catchments are currently generated for points at a 50m interval along the river network and this spacing was retained.

There is a need to explain 2 potentially confusing acronyms.

- The **IHDTM** is the Centre for Ecology and Hydrology (CEH) digital terrain model (DTM). Originally standing for the Institute of Hydrology DTM, the acronym has been retained to now refer to the Integrated Hydrological DTM. The grid size is 50m. It was derived from OS contours, spot heights, heightened lake shores, high water lines and digitised rivers from OS 1:50,000 maps (Morris and Heerdegen 1998).



- The **IHM** is the Environment Agency's Integrated Height Model 2014. The grid size is 2m and the vertical accuracy much better than older products such as the IHDTM. It consists of 72% Environment Agency LiDAR data, plus 2 other datasets to fill in the gaps.

In the late 1980s, CEH developed a bespoke code for creating drainage grids. The processes were developed specifically for application in a UK context at 50m scale based on the data sources available, at a time when there were no available off-the-shelf products for carrying out such tasks. Several specific algorithms were developed for resolving issues particular to this context. All FEH catchment boundaries and descriptors were based on this code. The IHDTM consists of 5 grids – elevation, surface type, outflow direction, inflow directions and cumulative catchment area. It covers the UK and the areas of the Republic of Ireland that drain into Northern Ireland.

The principal stages in the production of the IHDTM were as follows (Morris and Heerdegen 1998).

1. The heights of the digital rivers were calculated by non-linear interpolation between their intersections with contours, lake shores and coastline. Owing to positional inconsistencies between some rivers and contours, a small proportion of the rivers were not heightened.
2. At every point on a 50m square grid, interpolation between the heightened rivers and the above-listed OS datasets (treating the high water line as a 3m contour) was carried out to obtain an elevation grid that was spatially consistent with the river network. The method used took a weighted average of elevation estimates obtained from multiple transects that intersect at the point, with the weight for each estimate being based on the confidence in the interpolation.
3. A byproduct from the elevation gridding algorithm is the surface type grid. The value at each point is the most significant hydrological type in the 50m × 50m square centred on the point from – in increasing order of significance – land, river, lake and sea. Thus, a grid point may lie on land, but be classified as sea or lake so as to ensure the continuity of narrow (subgrid interval) estuaries and lakes within their gridded representation. Rivers in this grid are sometimes missing because the method uses heightened rivers and, as noted in (1) above, not all rivers were heightened. For a more comprehensive grid representation of the rivers, the outflow grid should be used – see (4) below.
4. The outflow grid was derived by: (a) representing the digital rivers as a continuous sequence of vectors on a 50m square grid; (b) in each lake, setting the directions at all the contained points so that they lead to the lake outlet; and (c) setting the direction at all other land points by reference to their elevation and the elevations of their 8 near neighbours. The algorithm used was designed to give a sequence of flow directions that follow the true plan direction of steepest slope and therefore did not set every individual point to flow to the neighbour that gave the steepest descent, as this would have led to sequences of points artificially following one of the 8 cardinal directions. Finally, where necessary, flow was directed uphill to provide continuity from local low points.

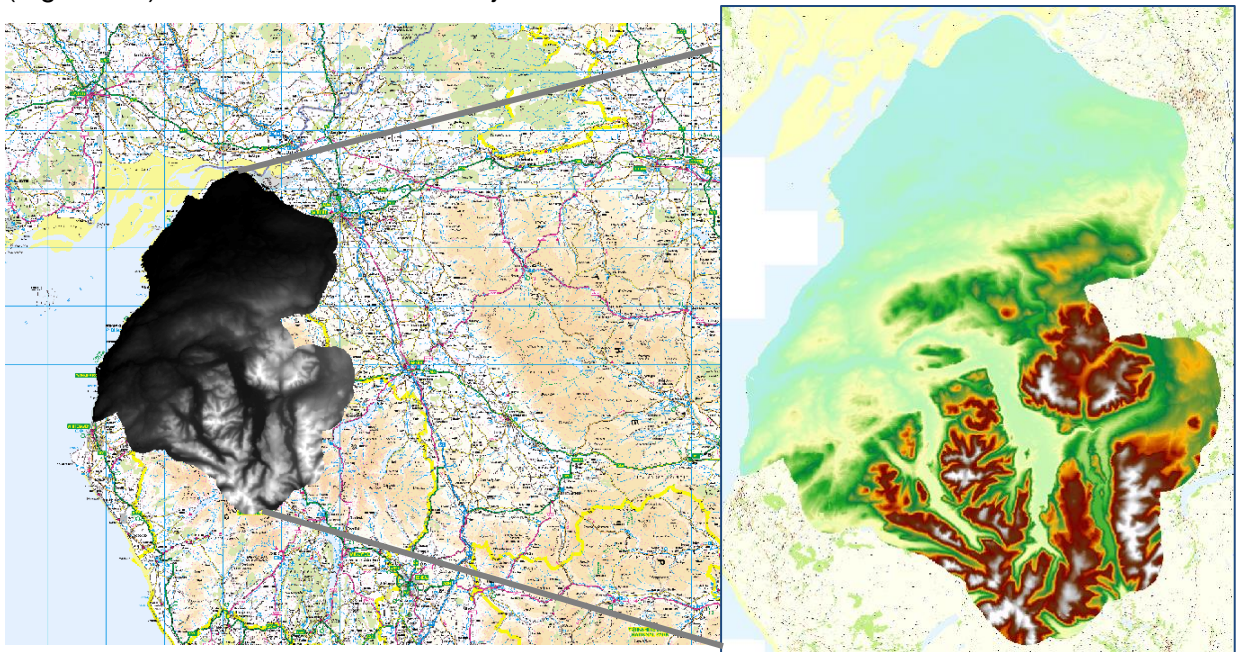
Within the present project, commercial software was investigated as an alternative to modifying the CEH code to run with high-resolution datasets. ArcMap, QGIS (Quantum GIS) and SCALGO (SCALable aLGOrithmics) were all investigated.

QGIS, with the addition of some external plugins, contains many of the tools needed to generate catchments. Many of the relevant tools within QGIS have their origins in GRASS (Geographic Resources Analysis Support System), another open-source GIS software package. With regards to SCALGO, the algorithms implemented allow SCALGO to handle very large datasets more efficiently than other GIS. However, the

licencing costs involved in obtaining the software meant it could not be utilised within the scope of this project beyond some exploratory analysis. ArcMap, with the additional ArcHydro toolbar, allows GIS users to generate all the datasets required in a consistent manner. One tool in the toolbar leads into the next, with expected column names and formats present, allowing processing to be easily carried out.

Some QGIS tools could be used in place of the ArcHydro tools, but consideration would have to be given to any additional processing required. For example, the flow direction raster in ArcHydro is a classified grid of 8 direction values (values of 1, 2, 4, 8, 16 and so on) representing the 8 flow directions possible from any given cell. The flow direction tool in QGIS also produces a flow direction raster, but contains the values from 0 to 360. The QGIS raster would need classifying if it were to be used in the ArcHydro tools for further tasks. In trying to avoid a mix of software, ArcMap (with ArcHydro) was initially deemed the most suitable software to use for processing.

Processing was carried out on a hydrometric area (HA) basis, concentrating largely on HA75, in north Cumbria, for development of the methods. This is a useful demonstration area as it includes both steep topography in the northern fringes of the Lake District and also the relatively flat coastal plain on the fringes of the Solway Firth (Figure 5.1). There are also some major lakes and reservoirs.



**Figure 5.1** Location of hydrometric area 75 in north Cumbria

### **5.2.1 Issues**

In developing the FEH methodology, the rivers were ‘burned’ into the DTM using a vector dataset since the vectors represented the river features more accurately than the coarse 50m DTM could. Imposing vector features onto the 2m DTM in HA75 has caused issues where there has been a mismatch between the path taken by the vector feature and the DTM representation.

With regards to the datasets available, there is no clear answer as to which data source (vector or raster) was more accurate and should be used as the starting point in the analysis. However, the availability of MasterMap data may cause a re-think in this respect. Burning rivers into the DTM caused issues in upland catchments; however, it may be deemed necessary in lowland catchments to impose some form of flow across the DTM, though large changes to the 2m DTM are not particularly desirable. Whichever process is adopted, a consistent approach across the entire country is vital.

The Environment Agency is currently carrying out a project to define probable overland flow paths and small catchments that flow into the river network for diffuse pollution and soil erosion models (personal communication from Alastair Duncan). For that work, the Environment Agency's Digital River Network has been burnt into the 2m resolution IHM, although only for a subset of rivers (river types 1 to 4 – primary, secondary, tertiary and lake/reservoir, and only those with primary flow direction). The burn depth varies according to river type. The intention is to introduce some of the culverted sections of the Digital River Network in locations where 'ponding' has occurred. Some of the processes developed during that investigation may be applicable to any future work on FEH catchment descriptors.

In generating catchment polygons every 50m along the river network, the limits of processing in ArcMap appear to have been reached. Multiple attempts to generate polygons ended with a memory corruption error. Approximately 2 weeks of processing yielded in the region of 50,000 polygons, out of a total of over 90,000 points generated at 50m spacing. The issue appears to be the larger catchments requiring more random access memory (RAM) than ArcMap can successfully manage, causing the processing to crash. The batch watershed creation can be resumed from the point where the crash occurred, but the memory corruption is a serious issue with regards to processing the DTM at a resolution of 2m.

ArcMap is not alone in having issues with handling the 2m DTM; both QGIS and SCALGO were unable to reliably handle processing the dataset. To move the project forward, the decision was taken to re-sample the DTM to a 10m resolution to reduce the computational load when processing the DTM data in ArcHydro. Once the DTM had been re-sampled, the subsequent processing steps were able to be reliably completed and with far shorter run times than when using the original 2m DTM.

The flow direction raster produced by ArcHydro exhibits a characteristic whereby large areas of a single flow direction appear in the raster. These large areas of a single value then affect the subsequent catchment delineation processing, as the catchment polygons cross these areas with an unnatural-looking straight line. The CEH process describes the application of a nearest neighbour filter to avoid this issue. Such an option is not available within the ArcHydro tools, nor with the SCALGO<sup>7</sup> or QGIS software packages. As described above, QGIS can produce a flow direction raster containing values from 1 to 360 degrees. But without applying a form of nearest neighbour analysis to smooth the outputs, water can still only flow in 1 of 8 directions from cell to cell, producing the straight line issues.

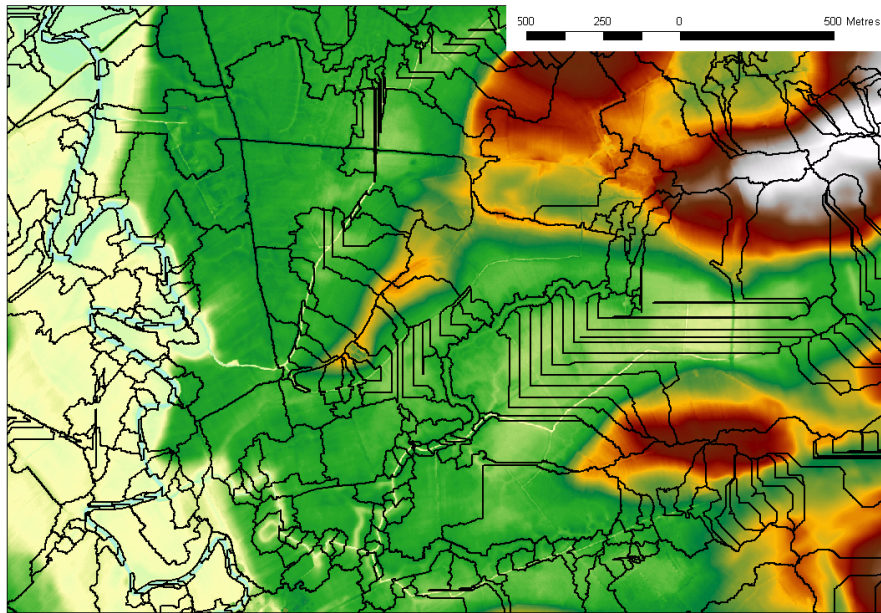
In Figure 5.2 and Figure 5.3 (centred around 325000, 556700), the straight lines of the generated catchment boundaries (in black) can clearly be seen cutting across the areas of a single colour. Figure 5.2 shows the DTM, with elevations ranging from 5 to 35m, while Figure 5.3 shows the flow direction raster.

Some concluding remarks on the development of a high-resolution drainage grid are given at the end of the following section, where catchment boundaries are compared with their equivalents from the FEH.

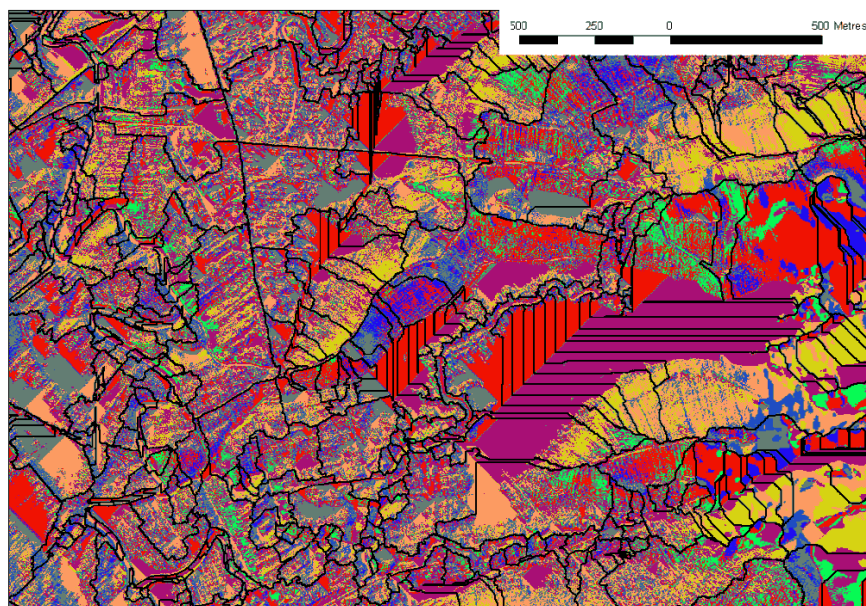
---

<sup>7</sup> When creating a flow direction grid, SCALGO does have an option called Aspect Decomposition, in which flow directions are assigned to 1 or 2 downslope cardinal direction cells depending on the local terrain aspect angle. However, it was not possible to create catchment boundaries within SCALGO when this option had been selected for the flow direction grid.





**Figure 5.2** DTM with portions of some catchment boundaries shown as straight lines in areas of uniform flow direction



**Figure 5.3** Flow direction grid with some catchment boundaries shown as straight lines in areas of uniform flow direction

## 5.3 Existing FEH catchment descriptors at high resolution

The aim of this task was to explore the feasibility of developing high-resolution catchment descriptors – first in combination with existing source datasets and, secondly, using improved source datasets such as higher resolution or more up-to-date thematic mapping.

The descriptors needed by the FEH and ReFH methods are listed in Table 5.1.

**Table 5.1 Catchment descriptors used by the FEH and ReFH methods**

| <b>Descriptor</b> | <b>Definition</b>  | <b>Derived from</b>   | <b>Used by which FEH methods</b> |
|-------------------|--|---|----------------------------------|
| AREA              | Area (km <sup>2</sup> )                                      | Catchment boundary  | Statistical and ReFH             |
| SAAR              | Annual average rainfall (mm)                                 | Catchment boundary + gridded rainfall                         | Statistical and ReFH             |
| BFIHOST           | Baseflow index estimated from Hydrology Of Soil Types (HOST) | Catchment boundary + gridded HOST classes                     | Statistical and ReFH             |
| FARL              | Flood attenuation due to reservoir and lakes                 | Catchment boundary + reservoir and lake shoreline vector data | Statistical                      |
| FPEXT             | Proportion of catchment covered in floodplains               | Catchment boundary + flood zone map                           | Statistical                      |
| DPLBAR            | Mean drainage path length (km)                               | Flow path network   | ReFH                             |
| DPSBAR            | Mean drainage path slope (m/km)                              | Flow path network and DTM                                     | ReFH                             |
| PROPWET           | Proportion of time catchment is wet                          | Catchment boundary + soil moisture data                       | ReFH                             |
| URBEXT2000        | Urban extent   | Catchment boundary + land cover data                          | Statistical                      |
| URBEXT1990        | Urban extent   | Catchment boundary + land cover data                          | ReFH                             |

The ReFH method also uses rainfall depth–duration–frequency information, but there are no plans to produce high-resolution versions of that dataset.

To automate the calculation of various catchment descriptors, an ArcMap AddIn was written in VB.NET. Using an AddIn gives better performance than VBA and also allows for better memory management, critical when carrying out large amounts of processing in ArcMap. For the purposes of this project, the code was written and executed without any form of graphical user interface (GUI) or user interaction required.

### 5.3.1 AREA

The area of a catchment polygon can easily be calculated using standard ArcMap functionality.

### 5.3.2 SAAR, BFIHOST, URBEXT, PROPWET

These were calculated using existing tools in ArcMap that calculate the mean value of datasets enclosed by polygons. SAAR data were provided as a dataset of points spaced on a regular 1km grid. HOST data were provided as a vector dataset of 1km grid squares. URBEXT2000 was calculated from the Land Cover Map 2000, using 3 classes of land cover – urban, suburban, and inland bare ground when within an urban settlement.

URBEXT1990 and PROPWET were not calculated, since neither was necessary for creating  $Q_T$  grids.

### 5.3.3 FARL

FARL is a measure of the flood attenuation by reservoirs and lakes for a given catchment polygon. It is calculated by obtaining the catchment area of each online lake, the area of the lake and the area of the catchment polygon. This is done using the attributes of the datasets and involves very little use of ArcObjects coding, relying more on the input datasets having had their attributes appropriately calculated before being processed. For the purpose of this pilot study, freely available OS datasets were used to define the extents of lakes and reservoirs.

### 5.3.4 FPEXT

FPEXT is a measure of floodplain extent. In the current FEH dataset, it is derived using the IH Report 130 floodplain map.<sup>8</sup> For the purposes of the present project, this was replaced with Flood Zone 3 (FZ3), taken from the current Flood Map.

Generating a value for this catchment descriptor involves pure ArcObjects code; no geoprocessing tools are required for this stage. For a given catchment polygon, the extent of the polygon is topologically compared with the FZ3 dataset to obtain a percentage coverage.

To process a large, topologically complex dataset such as FZ3, it is necessary to pre-process the dataset to reduce the size of any polygon being used by the ArcObjects code. The FZ3 dataset is therefore 'cookie cut' against a regular grid dataset to reduce any single polygon to a maximum of 200m × 200m in size. This processed dataset allows the ArcObjects code to better manage the calculations involved to the point where the code would not be able to complete successfully without the pre-processed dataset. Using the 'cookie cut' data does not alter the coverage or results of this processing.

The above code and processes can be reused to calculate the coverage of any polygon dataset against the catchment polygons.

---

<sup>8</sup> [www.ceh.ac.uk/services/ih130-digital-flood-risk-maps](http://www.ceh.ac.uk/services/ih130-digital-flood-risk-maps)

### 5.3.5 DPLBAR

DPLBAR is the mean drainage path length for a given catchment polygon. The catchment polygon and a flow direction raster are required to calculate this value. The AddIn code uses a built-in geoprocessing command to calculate the upstream distance along the flow paths of the input raster. The tool generates a raster output from which the mean value is extracted, using another of the built-in geoprocessing tools to obtain the raster property. The geoprocessing tools allow for better management of any errors or issues when calculating values than some of the ArcObjects objects and have been used wherever possible in the coding.

### 5.3.6 DPSBAR

DPSBAR is the mean drainage path slope value for a given catchment. The method for calculating DPSBAR involves the DTM and the flow direction raster datasets, along with the catchment polygon itself. The AddIn code clips the input raster datasets to the extent of the polygon being used. Clipping the rasters means the amount of data held in memory at any given time is reduced compared to holding the entire datasets in memory during the processing. Depending on the extent of the rasters involved, the datasets are then either processed in a single transaction or split into smaller datasets (based on a  $4 \times 4$  grid) before being processed.

In processing the datasets to calculate DPSBAR, the value of the DTM and the flow direction raster are obtained for each cell in the dataset. The value returned from the flow direction raster (1, 2, 4, 8, 16, 32, 64 or 128) then determines the direction that the code moves to find the appropriate cell in the DTM to read the elevation from so as to calculate the slope involved. Once the elevations involved have been obtained, the calculation of slope is then carried out.

### 5.3.7 Checks against FEH descriptors

FEH descriptors were obtained via the NRFA website for 11 gauging stations located in hydrometric area (HA) 75 in north Cumbria (see Figure 5.1). The appropriate catchment polygons for these locations were then extracted from the dataset generated by ArcHydro and the catchment descriptors calculated for comparison purposes. Any large differences in the catchment descriptor values would potentially indicate issues with the processes involved.

A summary of the catchment descriptors obtained from the website and those calculated using the catchment boundaries created by the ArcHydro processes is given in Table 5.2.

**Table 5.2 Comparison of FEH and high-resolution catchment descriptors at selected gauging stations**

| Station ID | Station name                         | Existing FEH descriptors |               |           |      | New descriptors from ArcHydro |               |           |      |
|------------|--------------------------------------|--------------------------|---------------|-----------|------|-------------------------------|---------------|-----------|------|
|            |                                      | Area (km <sup>2</sup> )  | DPSBAR (m/km) | SAAR (mm) | FARL | Area (km <sup>2</sup> )       | DPSBAR (m/km) | SAAR (mm) | FARL |
| 75001      | St Johns Beck at Thirlmere Reservoir | 42.1                     | 285           | 2611      | 0.72 | 30.9                          | 296           | 2788      | 0.69 |
| 75002      | Derwent at Camerton                  | 663                      | 208           | 1811      | 0.84 | 660.9                         | 231           | 1835      | 0.86 |
| 75003      | Derwent at Ouse Bridge               | 363                      | 247           | 2077      | 0.79 | 360.4                         | 274           | 2106      | 0.80 |
| 75004      | Cocker at Southwaite Bridge          | 116.6                    | 289           | 1975      | 0.83 | 116.2                         | 313           | 1996      | 0.85 |
| 75005      | Derwent at Portinscale               | 235                      | 245           | 2247      | 0.85 | 234.5                         | 275           | 2282      | 0.85 |
| 75006      | Newlands Beck at Braithwaite         | 33.9                     | 381           | 2390      | 0.99 | n/a                           | n/a           | n/a       | n/a  |
| 75007      | Glenderamackin at Threlkeld          | 64.5                     | 180           | 1733      | 0.99 | 61.5                          | 189           | 1774      | 1    |
| 75009      | Greta at Low Breiry                  | 145.6                    | 227           | 2039      | 0.91 | 144.7                         | 251           | 2056      | 0.92 |
| 75010      | Marron at Ullock                     | 26.8                     | 122           | 1508      | 0.96 | 28.1                          | 139           | 1478      | 0.97 |
| 75016      | Cocker at Scalehill                  | 64                       | 331           | 2252      | 0.71 | 63.3                          | 358           | 2276      | 0.74 |
| 75017      | Ellen at Bullgill                    | 96                       | 78            | 1110      | 0.98 | 103.3                         | 89            | 1109      | 0.99 |

At 6 of the 11 stations, the catchment area calculated from ArcHydro is within 1% of the FEH catchment area; at 8 of the 11, it is within 5%. Where the area matches within 5%, the SAAR matches to within 2% and the FARL to within 4%. There is a tendency for DPSBAR to be higher than the FEH values, by up to 14%. This may be because flow paths extend further up into the headwaters, with a minimum catchment area of 0.2km<sup>2</sup>.

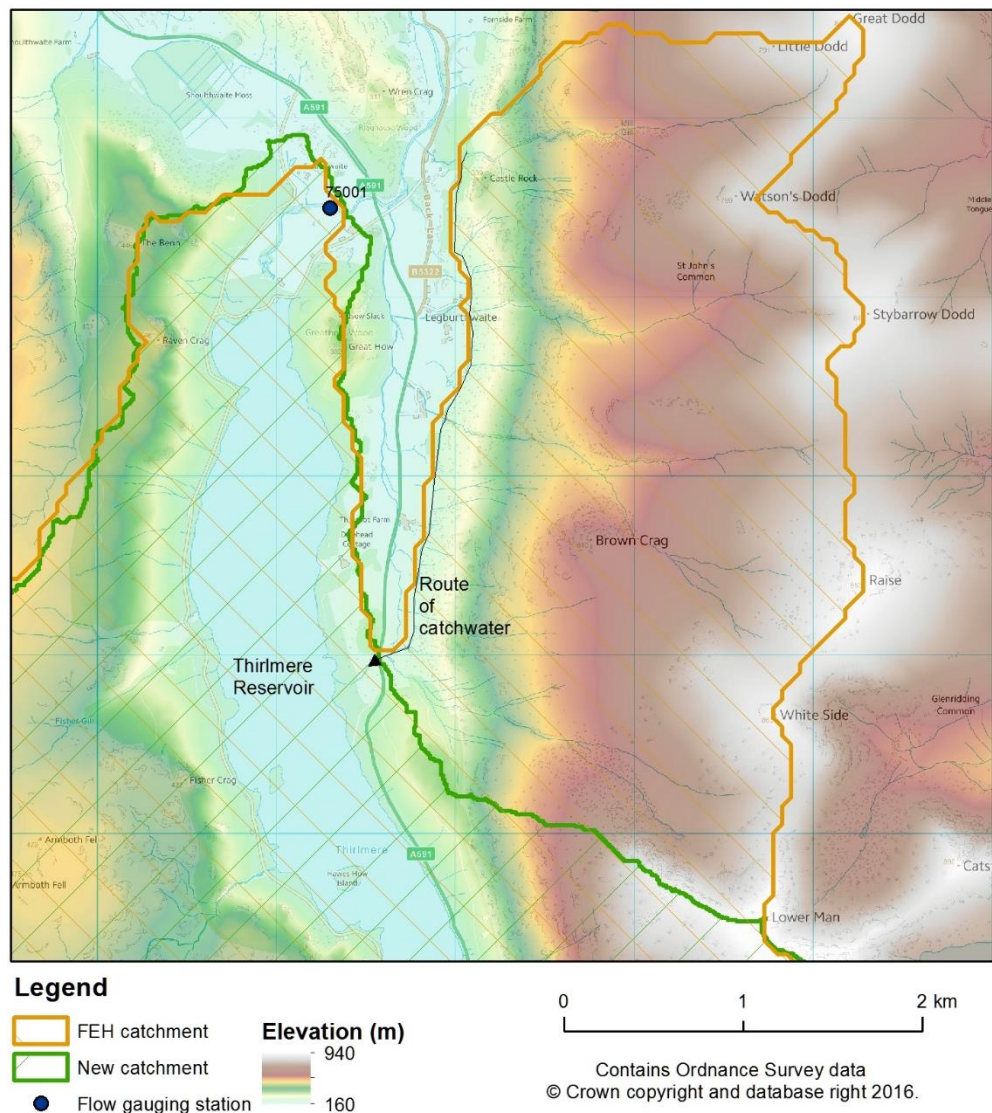
A variety of issues were encountered when comparing catchment boundaries for a number of the stations – 75001, 75002, 75006, 75007 and 75017. The typical issues are described below.

### *Gauging station 75001*

The 27% difference in catchment area is largely due to the presence of a man-made catchwater that diverts some streams on the western slope of the Helvellyn ridge in order to bring extra water into Thirlmere Reservoir (Figure 5.4). The FEH flow direction grid may have been altered to take account of the existence of the catchwater. The FEH catchment boundary is therefore more likely to be correct, at least for average flow conditions. In extreme floods, the capacity of the flow diversion may be exceeded, in which case it may be more appropriate to use the natural catchment boundary.

Comment: There are a number of instances of catchment boundaries that vary according to hydrological conditions. Others are seen in fenland areas. This underlines the importance of checking catchment boundaries with the benefit of local knowledge.



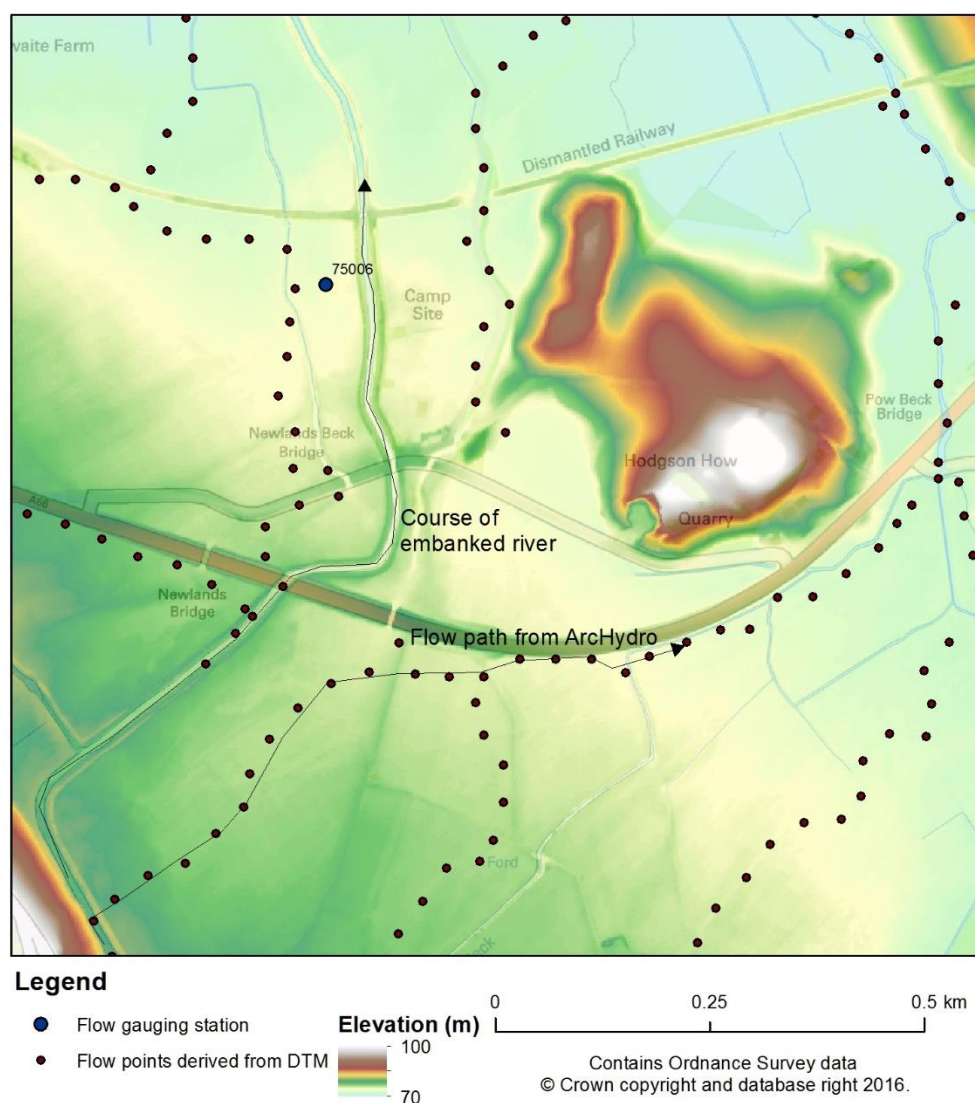


**Figure 5.4 Reason for catchment boundary differences for station 75001**

### *Gauging station 75006*

No comparable catchment boundary was generated using the ArchHydro processes. The reason is that a watercourse in the catchment, Newlands Beck, is not accurately represented by the flow points generated by ArchHydro. Towards the downstream end of the beck, the watercourse is constrained by embankments. These feature in the DTM and therefore affect the flow direction and flow accumulation processes, preventing the watercourse from being accurately represented.

Figure 5.5 shows how the flow path generated by ArchHydro deviates from the position of the river 1km upstream of the gauge, appearing to follow the natural topography rather than being constrained by the embankments. This is probably an example of where burning a river network into the DTM would have helped.

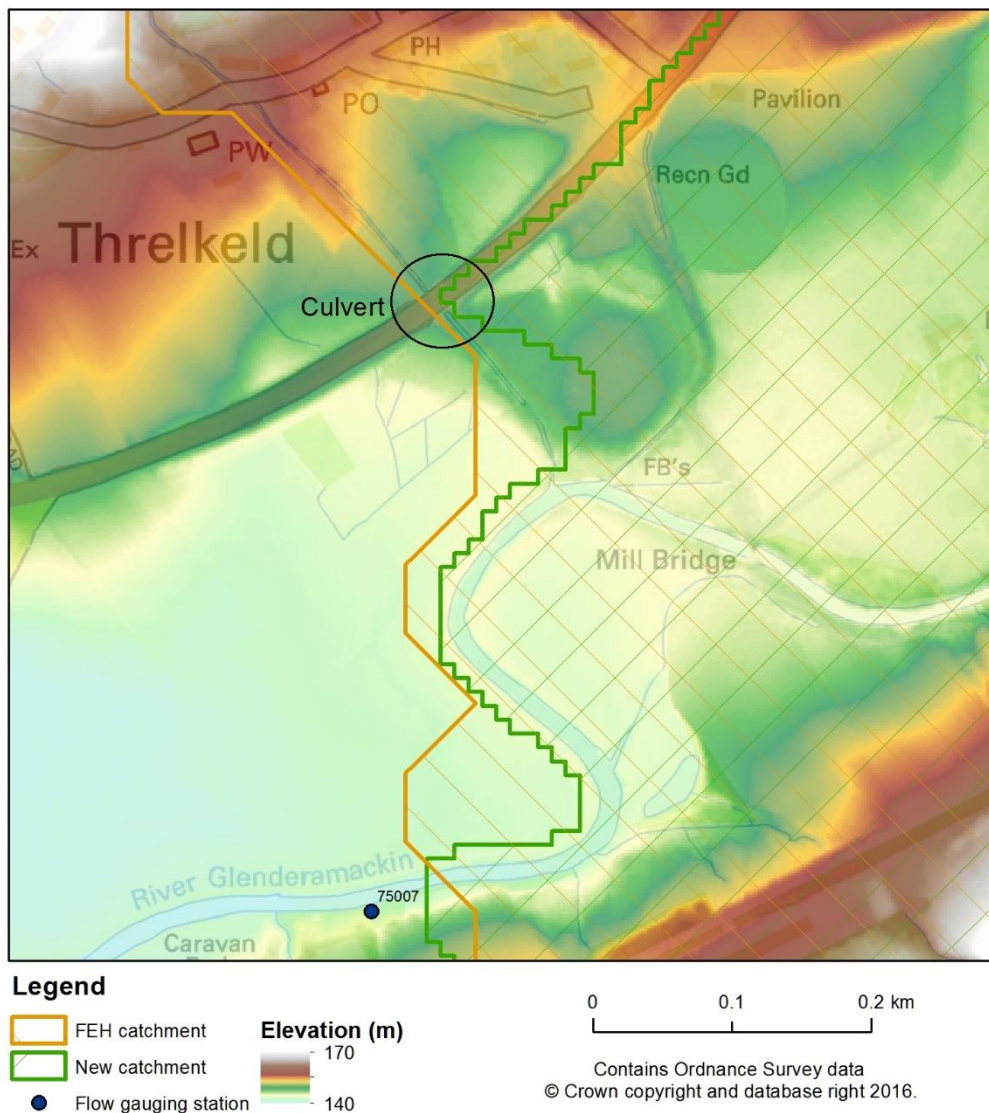


**Figure 5.5 Reason for catchment boundary differences for station 75006**

### *Gauging station 75007*

A road embankment on the A66 near Threlkeld affects the flow path representing Kilnhow Beck, with the culvert not represented in the DTM. As a result, the ArchHydro-derived catchment polygon is smaller than the FEH version, with the flows to the north of Threlkeld not entering the main watercourse (the Glenderamackin) until after the gauge location. This area is shown in Figure 5.6.

Comment: One drawback of a higher resolution DTM is that it is more likely to represent narrow topographic features such as road embankments. It is often necessary to edit the DTM to represent the presence of culverts, although it should be recognised that, in extreme floods or when the culvert inlet is blocked, flow may follow another overland route.



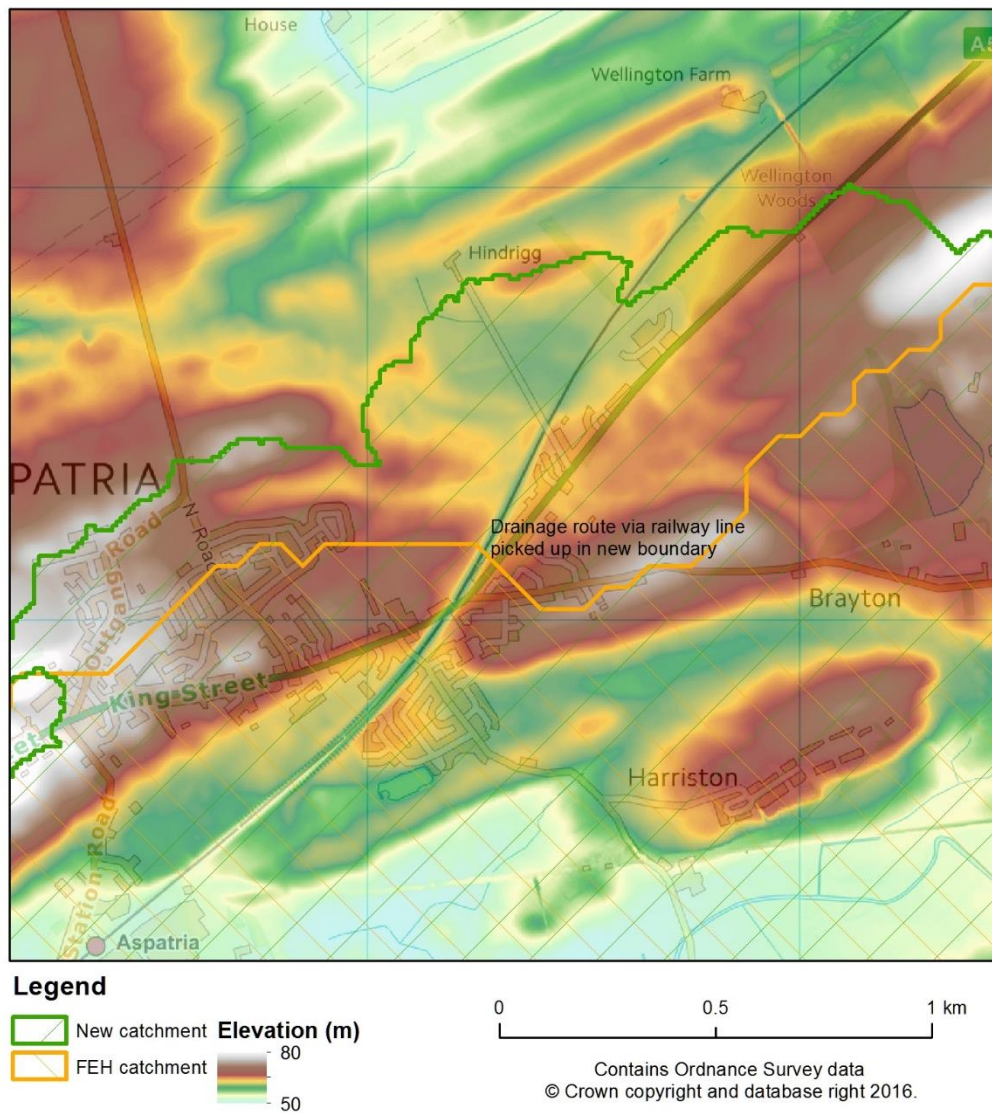
**Figure 5.6 Reason for catchment boundary differences for station 75007**

### *Gauging station 75017*

Another issue with using a more detailed DTM is that flow paths can be detected in locations where there are no watercourses. This can be seen in this case where the ArchHydro catchment polygon covers an additional area to the north of Aspatria, where a flow path has been detected flowing along the route of the rail line. This is shown in Figure 5.7, with the rail line running from the top right to the bottom left of the figure. The additional area accounts for the majority of the 7.3km<sup>2</sup> difference in catchment areas between the polygons in the 2 datasets.

Comment: From examining the DTM and map, it appears that this may be a genuine route for flood flow, even if, during normal conditions, water is kept off the railway via track drainage. However, a site inspection may not necessarily support this conclusion.





**Figure 5.7 Reason for catchment boundary differences for station 75017**

### Summary

The use of a more detailed DTM introduces new issues in the processes of defining catchment boundaries. Some of these issues are not apparent until a large number of lengthy processes have been carried out. They also rely on being able to compare the newly generated ArchHydro catchments against catchment boundaries derived from other methods. To resolve the issues related to culverted watercourses, a series of DTM edits could be made to allow the DTM to more accurately reflect the flow of water over the surface of the land. However, this is a significant task.

One option for processing on a national scale could be to generate a set of catchments initially for points at a wider spacing than 50m along the watercourse network such as 1km or more to generate a dataset for checking of flow paths. This would reduce the amount of abortive processing needed and the number of errors caused by incorrect flow paths. Once the checking and potential DTM edits had been carried out, the catchments could then be generated for points with a 50m spacing.

Where no DTM issues were encountered, it is encouraging to see very close comparisons between the datasets in terms of catchment area as well as catchment descriptor values calculated.

Processing the catchment descriptors for HA75 was a significant undertaking in terms of computing resources. The descriptors that require the interrogation of raster datasets, in particular DPLBAR and DPSBAR, involve conducting lengthy calculations. A modern PC with 16GB of RAM can only process somewhere in the order of 100 catchments every 45 minutes. The issue of badly managed RAM within ArcMap also affects the ability to run the processing over long periods of time, with ArcMap frequently crashing when it can no longer manage the RAM involved in processing. Applying the same processes at a national scale would be challenging. It may be that tools such as Python could be used for some of the processing stages to help reduce the impact of memory management in ArcMap.

## 5.4 Proposed new catchment descriptors

The existing FEH catchment descriptors were developed during the mid to late 1990s, at a time when digital datasets were relatively new. The calculation of the catchment descriptors for each of the 4 million UK catchments of at least 0.5km<sup>2</sup> in area used bespoke computer programs built around the IHDTM. However, even though the source data were often available at relatively coarse resolution, the catchment descriptors represented a huge advance on the maps published with the Flood Studies Report in 1975.

As digital datasets are now widely available at much higher resolution, and GIS makes the data processing much more tractable, it is appropriate to consider the options for developing new catchment descriptors from high-resolution source datasets. Derivation of the key FEH catchment descriptors required for the application of the FEH statistical method was described in the previous section for several test catchments. The feasibility of developing new catchment descriptors is discussed below.

In order for new catchment descriptors to be useful in flood frequency estimation, it will be necessary to recalibrate the FEH methods, an activity that remains outside the scope of this project. As a preliminary step, it would be necessary to calculate the descriptors across the whole of the UK, which would present immediate problems, since some datasets cover only England or England and Wales.

### 5.4.1 URBEXT2007/URBEXT2015

URBEXT2000 could be updated to URBEXT2007 using the Land Cover Map 2007 (LCM2007), available via the CEH website. As well as being more up-to-date, LCM2007 was derived from a wider range of data sources than the earlier LCM2000, including detailed field boundaries from OS MasterMap data, with 7 knowledge-based enhancement rules applied to improve the classification of land cover. The spatial resolution of the 2 datasets is identical (vector or 25m raster) and URBEXT2007 should be straightforward to calculate.

CEH is currently working on a further update called LCM2015, which is due for release in spring 2017.<sup>9</sup> This is expected to include some mapping of land cover change, making it possible in future to consider developing an index of catchment urbanisation through time for hydrological modelling.

### 5.4.2 PROPWET

The PROPWET catchment descriptor is currently required in the ReFH2 design hydrograph method for the estimation of model parameters at ungauged sites.

---

<sup>9</sup> [www.ceh.ac.uk/services/land-cover-map-2015](http://www.ceh.ac.uk/services/land-cover-map-2015)

PROPWET measures the proportion of time catchment soils are wet (that is, with a soil moisture deficit (SMD) of  $\leq 6\text{mm}$ ). It was calculated using a relatively complex procedure based on daily estimates of SMD derived from the Met Office's 40km gridded Meteorological Office Rainfall and Evapotranspiration Calculation System (MORECS) product.

The low resolution of PROPWET means that replacing it will provide serious benefits. Soil moisture data from the Met Office's Surface Exchange Scheme (MOSES) model are an obvious candidate for generating a new soil moisture catchment descriptor. The offline MOSES model operates using the same spatial (40km) and temporal (1 day) resolution as MORECS. However, another version of MOSES (online MOSES/MOSES-PDM/UKPP) is used for flood forecasting at a 2km, 1-hour resolution.

Several ways in which PROPWET could be improved have been identified, including direct use of SMD data from MOSES and better definition of SMD by accounting for land cover, rather than assuming that all grid squares are covered entirely in grassland with soil of medium water availability, as with MORECS. MOSES-PDM classifies land according to 1 of 10 types, where each tile can contain several land types.

One issue to be aware of is that both MORECS and MOSES have been found to suffer from problems thought to be due to inconsistencies in meteorological datasets. The Environment Agency is currently carrying out work to create an improved dataset of potential evapotranspiration (PE). Since the ReFH method has moved away from using SMD as a measure of soil moisture, it may be worth considering creating a replacement index that uses a catchment-average PE as a surrogate index for soil moisture. This would take advantage of the Environment Agency's forthcoming recalculation of PE, and it might also act as a useful complement to other catchment descriptors, in particular SAAR and BFIHOST.

Given that PROPWET is only concerned with an SMD of  $\leq 6\text{mm}$ , it may be possible to update it using satellite imagery. In this case, the variable of interest (SMD) would be measured directly rather than calculated from rainfall and soil data. Spatial resolution would also be greatly improved over its current 40km. However, the temporal resolution of the input data would be far lower than the 1-hour resolution that could be offered by MOSES-PDM, which may be significant in rapidly draining areas.

### **5.4.3 BFIHOST**

At the start of the project, there was an expectation that BFIHOST could be improved using higher resolution soils data. However, there does not appear to be an obvious alternative to the use of HOST data. Higher resolution soils maps (for example, NATMAP Vector) are not available throughout the UK, and the HOST2 project, a proposed replacement for HOST, is not now expected to go ahead.

NATMAP HOST is an existing product in which the 297 NATMAP soil classes are grouped into the 29 HOST classes. As with NATMAP, data are not available for Scotland, Northern Ireland or the Isle of Man. The existence of the 297 class NATMAP dataset anyway provides a base from which the HOST classes could be refined to account for potential subtle differences that the current HOST system cannot – though in England and Wales only.

HORIZON Hydraulics is a dataset of detailed hydraulic properties for each NATMAP class, based on laboratory analyses. It may be possible to use this dataset to refine estimates of BFI for each HOST class or to re-map the subsurface layers in MOSES-PDM using NATMAP. Neither is feasible within the scope of this project.

#### 5.4.4 Alternative morphometric descriptors

Eight morphometric (that is, landform) catchment descriptors were calculated by Jafaar and Han (2012) for 20 catchments in south-west England. These descriptors were form factor, average slope, maximum relief, relief ratio, drainage density, stream frequency, bifurcation ratio and length of overland flow. Broadly, equivalents for average slope and relief ratio already exist in the FEH as DPSBAR and DPLBAR, while  $\text{AREA/DPLBAR}^2$  is equivalent to form factor. Maximum relief measures the elevation difference between the highest and lowest points of a catchment.

The last 4 in the list of morphometric catchment descriptors all relate to properties of the river network and no equivalents exist in the FEH.

- Drainage density measures the total stream length per catchment area.
- Length of overland flow is one-half of the reciprocal of this.
- Stream frequency measures the total number of stream segments per catchment area.
- Bifurcation ratio measures the number of  $n$ th order streams to  $(n+1)$ th order streams.

None of these river network catchment descriptors was found useful for QMED estimation in the 20 catchments for which they were derived. However, the study considered only a small number of catchments, all in the same region.

In contrast, drainage density was found to be a useful predictor of QMED in Ireland and is included in the Flood Studies Update procedures (Murphy et al. 2014). Furthermore, river network descriptors are easy to calculate and may have other uses, and thus may be worthy of further research.

### 5.5 Automatic estimation of design flows for all points along rivers

The project scoping document called for the production of a point dataset with estimated design flows for a range of return periods at evenly spaced points along rivers – an automated application of the FEH methods. This sort of dataset was first produced by CEH in 2003 (Morris 2003).

Design peak flows for the pilot area, HA75 in north Cumbria, were estimated using the current version of the FEH statistical method. The steps involved in the process were as follows.

1. Import catchment descriptors for each location.
2. Estimate QMED using the regression equation given in Environment Agency (2008).
3. Adjust QMED by automatic identification of a donor site, again using the procedures from Environment Agency (2008). A single donor site was identified automatically for every subject site, choosing the closest gauged catchment (according to distance between centroids) that is rural and has flood peak data suitable for estimating QMED. The adjustment factor was moderated according to the distance between the 2 catchments. The analysis was carried out using the NRFA peak flow database (version 3.3.4).
4. Adjust QMED for urbanisation using the procedure given by Kjeldsen (2010). The procedure for urban adjustment of QMED differs from that used in the

WINFAP-FEH v3 software. The difference between the 2 approaches is only significant on urbanised catchments that are also permeable.

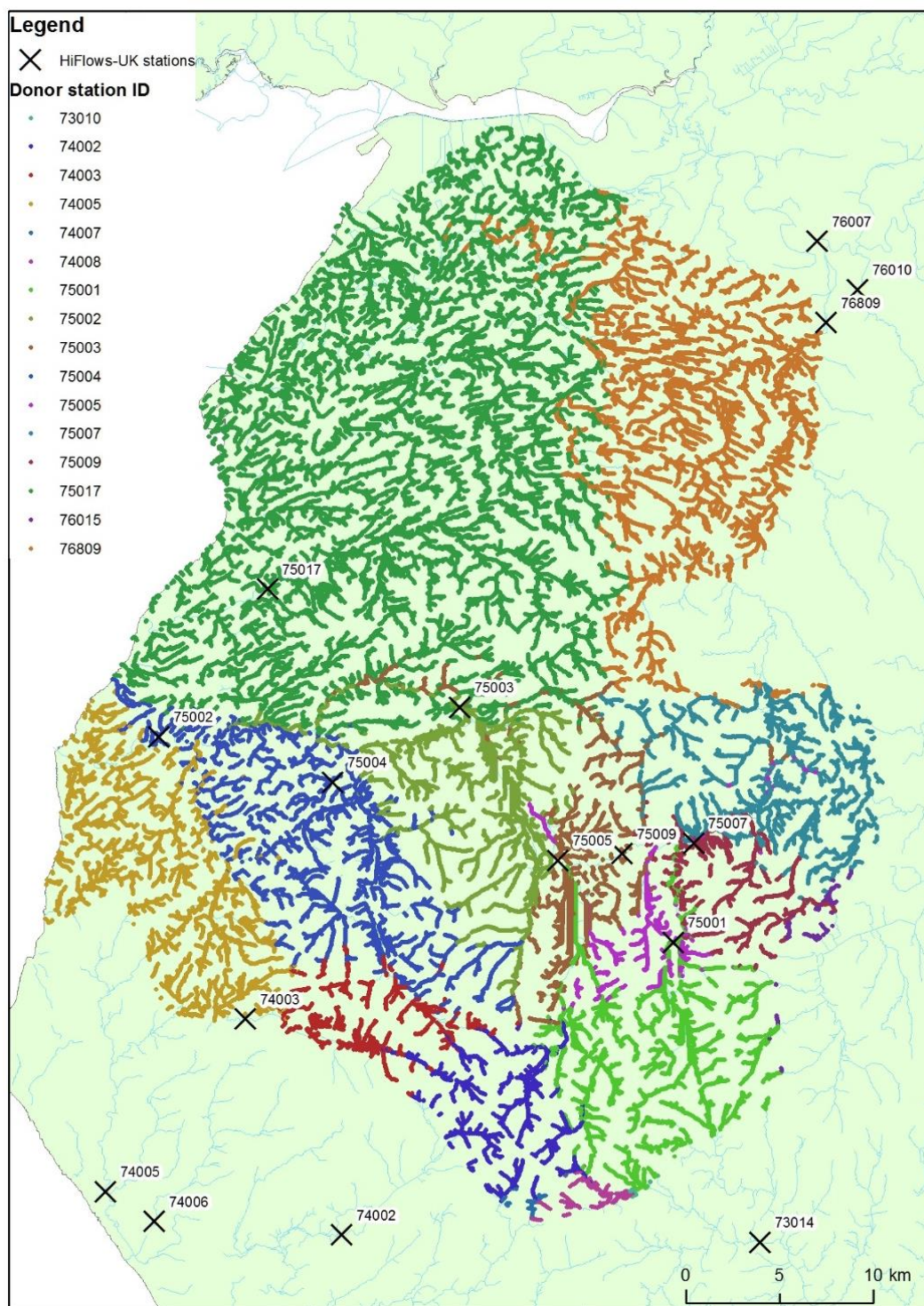
5. Construct a pooling group using the method given in Environment Agency (2008). A pooling group was developed for each flow estimation point using the NRFA peak flow database.
6. Develop a pooled growth curve using the methods described in Environment Agency (2008) to weight results from gauges in the pooling group. A GLO distribution was used to represent the growth curve. The growth curve was adjusted for urbanisation using the methods from Kjeldsen (2010).
7. Scale the growth curve by QMED to give the design flows for a range of return periods.

The calculations were performed using existing software, JFes, developed by JBA Consulting. It is a web-based application written in VB.NET.

The following maps illustrate the results. Figure 5.8 illustrates the choice of donor site at each location and Figure 5.9 maps the resulting adjustment factors for QMED. In many locations, adjustment factors are close to 1, owing to the long distance between the catchment centroid and that of the closest gauged catchment. The greatest adjustments, in the range 1.41 to 1.55, are found in the far south (Figure 5.9), where there is a small area for which the donor station is 74003 (Figure 5.8) – the River Ehen at Bleach Green.

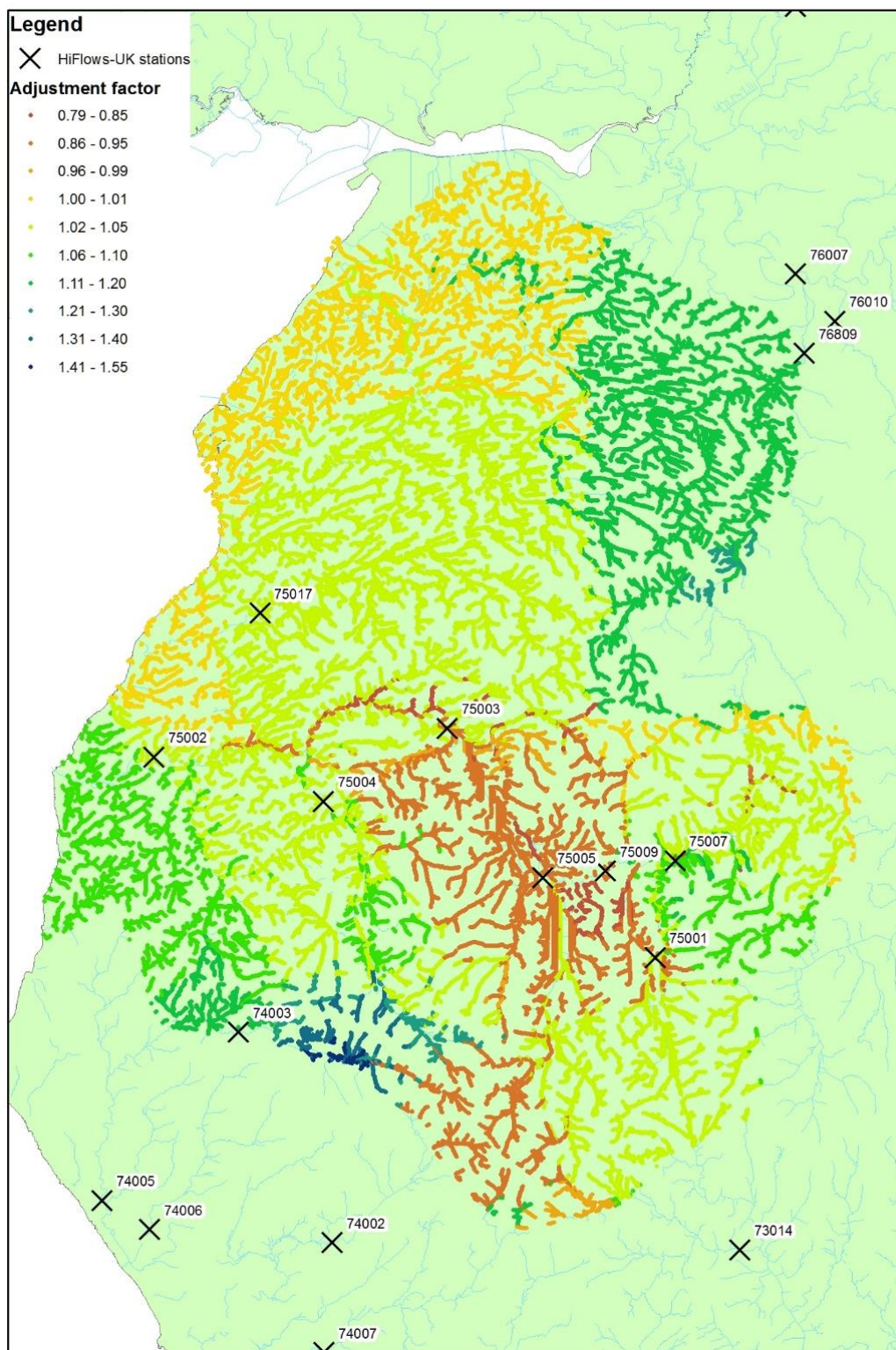
The spatial variation of the 100-year growth factors is shown in Figure 5.10. Growth factors are generally low in the southern mountainous area of HA75, where SAAR is high. Therefore, pooling groups are likely to contain high rainfall catchments that record large floods every year and so tend to exhibit shallow growth curves. Particularly low growth factors are seen downstream of lakes such as on the River Derwent, where low FARL values influence the composition of the pooling group. There are also some watercourses in the north with low growth factors, probably due to large values of FPEXT.





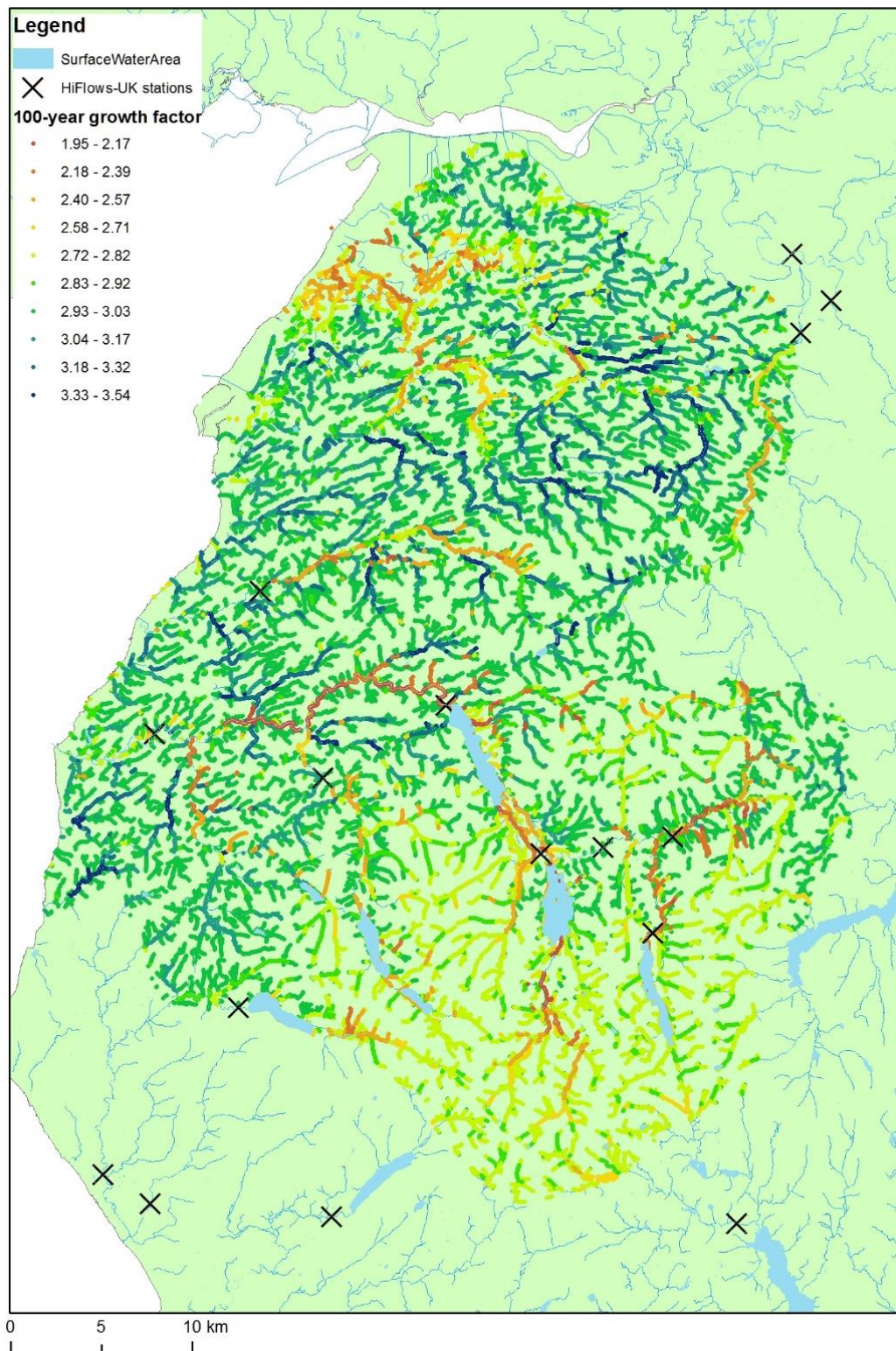
**Figure 5.8** Illustration of donor station used for adjusting QMED at each studied location in HA75





**Figure 5.9** Illustration of adjustment factor for QMED at each studied location in HA75





**Figure 5.10** Illustration of growth factor for the 100-year flood at each studied location in HA75

## 5.6 Closing comments on high-resolution catchment descriptors

There is potential to replace the FEH catchment descriptors with a new dataset calculated from contemporary high-resolution digital terrain data, along with more up-to-date, accurate and high-resolution spatial datasets such as for floodplain extent, urban extent and soil wetness. There is some demand for this sort of improvement, for example, in the urban runoff management sector. It could be valuable to be able to distinguish between different types of urban land cover, including perhaps an index that defined areas drained by sustainable drainage systems.

The benefits of high-resolution descriptors are difficult to judge. Until new descriptors have been developed and calculated for all gauged catchments in the UK, their benefit in terms of reducing uncertainty in flood frequency estimation are not known. The FEH research developed a large number of catchment descriptors, only some of which were found to be useful in estimating design flows for ungauged catchments. It may be that the largest benefits are seen on very small catchments, although these benefits may be difficult to measure given the relatively small number of such catchments for which flood peak data are available. Research into flood estimation on small catchments is being carried out within Environment Agency project SC090031, Estimating flood peaks and hydrographs for small catchments (Phase 2), and this includes an investigation of alternative soil descriptors.

This pilot study has shown that it is possible to develop high-resolution catchment descriptors. However, further work is required to:

- remove some of the problems encountered with catchment boundaries
- streamline the computational processes so that they can be scaled up nationally
- arrange licencing of spatial datasets covering the whole of the UK for commercial use
- develop new descriptors
- recalibrate the FEH procedures accordingly

Perhaps one conclusion to take from this investigation is that the UK already benefits from a high quality dataset of catchment descriptors that is suitable for most purposes.

## 6 Recommendations for further work

The primary recommendation arising from this project is to implement the routine use of local data in UK flood estimation in order to reduce uncertainty. Current practice falls far short of this ideal, despite the fact that guidance on the use of some types of local data has been available to practitioners for many years. There are many possible reasons for this including:

- commercial factors associated with the way flood hydrology is procured
- a greater appreciation of the costs than of the benefits of local data
- a lack of awareness, confidence, skills or software tools

This objective will not be met solely by research and the development and dissemination of guidance. These need to go hand-in-hand with other improvements, such as:

- educating practitioners and clients
- expanding project specifications to require more than the use of standard methods
- more rigorous review of flood estimates
- a change in culture that gets hydrologists out more often from their computer models into the field

The following recommendations for further research are offered.

There is a need for a method to quantify uncertainty within the enhanced single-site procedure when the FEH statistical method is applied at gauged sites. Research is needed to work out how to account for sampling error in these circumstances.

Within the ReFH2 method, there may be scope to introduce a data transfer procedure for the initial soil moisture,  $C_{ini}$ , similar to that for QMED in the FEH statistical method. This could help to reduce uncertainty in the results of ReFH2 at ungauged sites.

Research is needed to provide guidance on how to introduce historical flood data to a pooled analysis within the FEH statistical method. By extending the maximum likelihood approach, it may be possible to build models to obtain regional estimates in which the historical information of any site can be included in the analysis. A possible avenue is explored in Section B.8.

The investigation of bankfull channel width was limited by the small size of the dataset. It is recommended that any future update to the FEH statistical method considers including the compilation of a reliable dataset of channel geometry data for all gauging stations, so that measures of channel geometry can be included as candidate variables in the regression or an alternative regression. It would be worth exploring the utility of LiDAR data, along with aerial photography for the estimation of channel widths.

FEH Local has developed 2 case studies of how palaeoflood data can be included in flood frequency estimation; these are presented in the practitioner guidance. It is recommended that the uncertainty associated with the reconstruction of discharges from palaeoflood data is further investigated. It would be valuable to identify any examples where there is enough confidence in palaeoflood data to justify a convincing case for making a significant alteration to design floods estimated using FEH methods.

Finally, this report includes a proposal for development of a system to enable recording, quality assurance and widespread sharing of local flood data. Much of this information has a relevance that is much wider than flood frequency estimation. It is recommended that this archive is set up within the NRFA and populated as part of a broader effort to encourage the use of local data in flood frequency estimation.

# References

- ARCHER, D.R. AND FOWLER, H.J., 2015. Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain. *Journal of Flood Risk Management*, DOI: 10.1111/jfr3.12187.
- ARCHER, D., FOSTER, M., FAULKNER, D. AND MAWDSLEY, J., 2000. The synthesis of design flood hydrographs. In *Flooding: Risks and Reactions*, Proceedings of CIWEM/ICE conference (London, 5 October 2000), 45-57. London: Chartered Institution of Water and Environmental Management.
- AZZALINI, A., 1996. *Statistical Inference Based on the Likelihood*. Boca Raton, FL: CRC Press.
- BAYLISS, A.C. AND JONES, R.C., 1993. *Peaks-over-threshold flood database: summary statistics and seasonality*. Report No. 121. Wallingford: Institute of Hydrology.
- BAYLISS, A.C. AND REED, D.W., 2001. *The use of historical data in flood frequency estimation*. Wallingford: Centre for Ecology & Hydrology.
- BENSON, M A., 1950. Use of historical data in flood-frequency analysis. *Eos, Transactions of the American Geophysical Union*, 31 (3), 419-424.
- BEVEN, K., 2010. *Environmental Modelling: An Uncertain Future?* Boca Raton, FL: CRC Press.
- BEVEN, K. and HALL, J. (EDS), 2014. *Applied Uncertainty Analysis for Flood Risk Management*. London: Imperial College Press.
- BEVEN, K., LEEDAL, D., MCCARTHY, S., LAMB, R., HUNTER, N., BATES, P., NEAL, J. AND WICKS, J., 2014. *Framework for assessing uncertainty in fluvial flood risk mapping*. CIRIA Report C721. London: CIRIA.
- BEZAK, N., BRILLY, M. AND ŠRAJ, M., 2014. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, 59 (5), 959-977.
- BLACK, A.R. AND LAW, F.M., 2004. Development and utilization of a national web-based chronology of hydrological events. *Hydrological Sciences Journal*, 49 (2), 237-246.
- BLACK, A. AND FADIPE, D., 2009. Use of historic water level records for reassessing flood frequency: case study of the Spey catchment. *Water and Environment Journal*, 23 (1), 23-31.
- BOOKER, D.J., 2010. Predicting wetted width in any river at any discharge. *Earth Surface Processes and Landforms*, 35 (7), 828-841.
- BOOKER, D.J. AND DUNBAR, M.J., 2008. Predicting river width, depth and velocity at ungauged sites in England and Wales using multilevel models. *Hydrological Processes*, 22 (20), 4049-4057.
- BOORMAN, D.B., HOLLIS, J.M. AND LILLY, A., 1995. *Hydrology of Soil Types: a hydrologically-based classification of the soils of the United Kingdom*. IH report 126. Wallingford: Institute of Hydrology.
- BROWN, A.G., 2003. Global environmental change and the palaeohydrology of western Europe: a review. In *Palaeohydrology: Understanding Global Change* (ed. K.J. Gregory and G. Benito), Chapter 8, pp. 105-121. Chichester: John Wiley & Sons.

- BURN, D.H., 1990. An appraisal of the 'region of influence' approach to flood frequency analysis. *Hydrological Sciences Journal*, 35 (2), 149-165.
- CLARK, C., 2004. Estimating extreme floods: the case of the Great Till Flood of 1841. *International Water Power and Dam Construction*, 56, 14-19.
- CLARK, C., 2014. *Resistance to the use of historic floods in order to estimate extreme floods*. Abstract for British Hydrological Society 12th National Symposium, University of Birmingham, September 2014.
- COHN, T.A., LANE, W.L. AND BAIER, W.G., 1997. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resources Research*, 33 (9), 2089-2096.
- COLES, S.G., 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.
- DODOV, B. AND FOUFOULA-GEORGIOU, E., 2004. Generalized hydraulic geometry: derivation based on a multiscaling formalism. *Water Resources Research*, 40 (6), W06302.
- ENGINEERS AUSTRALIA, 2016. *Australian Rainfall & Runoff, 4th edition. Book 3. Peak flow estimation* [online]. Canberra: Commonwealth of Australia (Geoscience Australia). Available from: <http://arr.ga.gov.au> [accessed 19 October 2016].
- ENVIRONMENT AGENCY, 2008. *Improving the FEH statistical procedures for flood frequency estimation*. Science Report SC050050. Bristol: Environment Agency.
- ENVIRONMENT AGENCY, 2014. *Management of the London Basin Chalk aquifer: status report 2014*. Bristol: Environment Agency.
- FISHER, N.I., 1993. *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.
- FOULDS, S.A. AND MACKLIN, M.G., 2016. A hydrogeomorphic assessment of twenty-first century floods in the UK. *Earth Surface Processes and Landforms*, 41 (2), 256-270.
- FRANCES, F., SALAS, J.D. AND BOES, D.C., 1994. Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models. *Water Resources Research*, 30 (6), 1653-1664.
- GAÁL, L., SZOLGAY, J., KOHNOVÁ, S., HLAVČOVÁ, K. AND VIGLIONE, A., 2010. Inclusion of historical information in flood frequency analysis using a Bayesian MCMC technique: a case study for the power dam Orlík, Czech Republic. *Contributions to Geophysics and Geodesy*, 40 (2), 121-147.
- GAUME, E., GAÁL, L., VIGLIONE, A., SZOLGAY, J., KOHNOVÁ, S. AND BLÖSCHL, G., 2010. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 394 (1-2), 101-117.
- GUSTARD, A., BULLOCK, A. AND DIXON, J.M., 1992. *Low flow estimation in the United Kingdom*. IH report 108. Wallingford: Institute of Hydrology.
- HARVEY, A.M., 1969. Channel capacity and the adjustment of streams to hydrologic regime. *Journal of Hydrology*, 8 (1), 82-98.
- HAUGHTON, G., BANKOFF, G. AND COULTHARD, T.J., 2015. In search of 'lost' knowledge and outsourced expertise in flood risk management. *Transactions of the Institute of British Geographers*, 40 (3), 375-386.



- HOSKING, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52 (1), 105-124.
- HOSKING, J.R.M., 1995. The use of L-moments in the analysis of censored data. In *Recent Advances in Life-Testing and Reliability* (ed. N. Balakrishnan), pp. 545-564. Boca Raton, FL: CRC Press.
- HOSKING, J.R.M. AND WALLIS, J.R., 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge: Cambridge University Press.
- INSTITUTE OF HYDROLOGY, 1980. *Low flow studies report No 1*. Research Report. Wallingford: Institute of Hydrology.
- INSTITUTE OF HYDROLOGY, 1983. *Some suggestions for the use of local data in flood estimation*. Flood Studies Supplementary Report No. 13. Wallingford: Institute of Hydrology.
- INSTITUTE OF HYDROLOGY, 1999. *Flood Estimation Handbook* (5 volumes). Wallingford: Institute of Hydrology.
- JAAFAR, W.Z.W. AND HAN, D., 2012. Catchment characteristics for index flood regionalisation. *Water Management*, 165 (3), 179-189.
- JBA CONSULTING, 2009. *Severity assessments North East Region*. Report to the Environment Agency.
- JBA CONSULTING, 2010. *Review of hydrology for Five Oak Green Flood Alleviation Scheme*. Report to the Environment Agency.
- JBA CONSULTING, 2014. *2013–2014 post-event flood analysis*. Four Area Reports to Environment Agency, South East Region.
- JONES, A.F., MACKLIN, M.G. AND LEWIN, J., 2010. Flood series data for the Later Holocene: available approaches, potential and limitations from UK alluvial sediments. *The Holocene*, 20 (7), 1123-1135.
- JONES, A.F., MACKLIN, M.G. AND BREWER, P.A., 2012. A geochemical record of flooding on the upper River Severn, UK, during the last 3750 years. *Geomorphology*, 179, 89-105.
- JONG-JUNE, J., YOUNG-OH, K. AND YONGDAI, K., 2011. Expected probability weighted moment estimator for censored flood data. *Advances in Water Resources*, 34 (8), 933-945.
- KAPLAN-HENRY, T., 2007. McNally post-fire discharge and the relationship of Sierra Nevada-wide flood frequency curves and local Kern River discharge curves. In *Advancing the Fundamental Sciences: Proceedings of the Forest Service National Earth Sciences Conference* (San Diego, CA, 18–22 October 2004), edited by M. Furniss, C. Clifton and K. Ronnenberg, Volume 1, 27-35. General Technical Report PNW-GTR-689. Portland, OR: USDA Fire Service, Pacific Northwest Research Station.
- KJELDSEN, T.R., 2007. *Flood Estimation Handbook Supplementary Report No. 1. The Revitalised FSR/FEH rainfall-runoff method*. Wallingford: Centre for Ecology & Hydrology.
- KJELDSEN, T.R., 2010. Modelling the impact of urbanization on flood frequency relationships in the UK. *Hydrology Research*, 41 (5), 391-405.
- KJELDSEN, T.R., 2015. How reliable are design flood estimates in the UK? *Journal of Flood Risk Management*, 8 (3), 237-246.

- KJELDSEN, T.R. AND JONES, D.A., 2004. Sampling variance of flood quantiles from the generalised logistic distribution estimated using the method of L-moments. *Hydrology and Earth System Sciences*, 8 (2), 183-190.
- KJELDSEN, T.R. AND JONES, D.A., 2006. Prediction uncertainty in a median based index flood method using L-moments. *Water Resources Research*, 42, W07414.
- KJELDSEN, T.R. AND JONES, D., 2007. Estimation of an index flood using data transfer in the UK. *Hydrological Sciences Journal*, 52 (1), 86-98.
- KJELDSEN, T.R. AND PROSDOCIMI, I., 2015. A bivariate extension of the Hosking and Wallis goodness-of-fit measure for regional distributions. *Water Resources Research*, 51 (2), 896-907.
- KJELDSEN, T.R., MILLER, J.D. AND PACKMAN, J.C., 2013. Modelling design flood hydrographs in catchments with mixed urban and rural land cover. *Hydrology Research*, 44 (6), 1040-1057.
- KJELDSEN, T.J., JONES, D.A AND MORRIS, D.G., 2014a. Using multiple donor sites for enhanced flood estimation in ungauged catchments. *Water Resources Research*, 50 (8), 6646-6657.
- KJELDSEN, T.R., LAMB, R. AND BLASKOVA, S., 2014b. Uncertainty in flood frequency analysis. In *Applied Uncertainty Analysis for Flood Risk Management* (ed. K. Beven and J. Hall), pp. 153-197. London: Imperial College Press.
- KJELDSEN, T.R., MACDONALD, N., LANG, M., MEDIERO, L., ALBUQUERQUE, T., BOGDANOWICZ, E., BRAZDIL, R., CASTELLARIN, A., DAVID, V., FLEIG, A., GUL, G. O., KRIAUCIUNIENE, J., KOHNOVÁ, S., MERZ, B., NICHOLSON, O., ROALD, L. A., SALINAS, J. L., SARAUSKIENE, D., ŠRAJ, M., STRUPCZEWSKI, W., SZOLGAY, J., TOUMAZIS, A., VANNEUVILLE, W., VEIJALAINEN, N. AND WILSON, D., 2014c. Documentary evidence of past floods in Europe and their utility in flood frequency estimation. *Journal of Hydrology*, 517, 963-973.
- LAWLOR, S.M., 2004. *Determination of channel-morphology characteristics, bankfull discharge, and various design-peak discharges in western Montana*. USGS Scientific Investigations Report 2004-5263. Reston, VA: US Geological Survey.
- LEESE, M.N., 1973. Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. *Water Resources Research*, 9 (6), 1534-1542.
- LLOYD'S, 2012. *Geomorphology and changing flood risk in the UK*. London: Lloyd's Emerging Risks Team.
- LVOVITCH, M.I., 1972. Hydrologic budget of continents and estimate of the balance of global fresh water resources. *Soviet Hydrology*, 4, 349-360.
- MACDONALD, N., 2014. Millennial scale variability in high magnitude flooding across Britain. *HESS Discussions*, 11 (9), 10157-10178.
- MACDONALD, N. AND BLACK, A.R., 2010. Reassessment of flood frequency using historical information for the River Ouse at York, UK (1200-2000). *Hydrological Sciences Journal*, 55 (7), 1152-1162.
- MACDONALD, N., KJELDSEN, T.R., PROSDOCIMI, I. AND SANGSTER, H., 2014. Reassessing flood frequency for the Sussex Ouse, Lewes: the inclusion of historical flood information since AD 1650. *Natural Hazards and Earth System Sciences*, 14 (10), 2817-2828.

- MADSEN, H., RASMUSSEN, P.F. AND ROSBJERG, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At-site modelling. *Water Resources Research*, 33 (4), 747-757.
- MARSH, T.J. AND HANNAFORD, J. (EDS), 2008. *UK Hydrometric Register. Hydrological data UK series*. Wallingford: Centre for Ecology & Hydrology.
- MERZ, R. AND BLÖSCHL, G., 2008. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, 44 (8), W08432.
- MILLS, P., NICHOLSON, O. AND REED, D., 2014. *Flood studies update. Technical Research Report Volume IV: Physical catchment descriptors*. Ireland: Office of Public Works.
- MODRICK, T.M. AND GEORGAKAKOS, K.P., 2014. Regional bankfull geometry relationships for southern California mountain streams and hydrologic applications. *Geomorphology*, 221, 242-260.
- MORRIS, D.G., 2003. *Automation and appraisal of the FEH statistical procedures for flood frequency estimation*. Report to Defra. Wallingford: Centre for Ecology & Hydrology.
- MORRIS, D.G. AND HEERDEGEN, R.G., 1998. Automatically derived catchment boundaries and channel networks and their hydrological applications. *Geomorphology*, 1 (2), 131-141.
- MURPHY, C., CUNNANE, C., DAS, S. and MANDAL, U., 2014. *Flood Studies Update Technical Research Report Volume 11: Flood frequency estimation*. Dublin: Office of Public Works.
- NERC, 1975. *Flood Studies Report* (5 volumes). London: Natural Environment Research Council.
- NEPPEL, L., RENARD, B., LANG, M., AYRAL, P.-A., COEUR, D., GAUME, E., JACOB, N., PAYRASTRE, O., POBANZ, K. AND VINET, F., 2010. Flood frequency analysis using historical data: accounting for random and systematic errors. *Hydrological Sciences Journal*, 55 (2), 192-208.
- NGUYEN, C., GAUME, E. AND PAYRASTRE, O., 2014. Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, 508, 385-396.
- O'BRIEN, N.L. AND BURN, D.H., 2014. A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. *Journal of Hydrology*, 519, 2040-2048.
- O'CONNELL, D.R.H., LEVISH, D.R. AND OSTENAA, D.A., 1998. Risk-based hydrology: Bayesian flood-frequency analyses using paleoflood information and data uncertainties. In *Proceedings of the First Federal Interagency Hydrologic Modeling Conference* (Las Vegas, 19–23 April 1998), Volume 1, pp. 4-101-4-108. Reston, VA: Office of Water Quality, Subcommittee on Hydrology of the Interagency Advisory Committee on Water Data.
- PAINE, J.L., ROWAN, J.S. AND WERRITTY, A., 2002. Reconstructing historic floods using sediments from embanked flood plains: a case study of the River Tay in Scotland. In *The Structure, Function and Management Implications of Fluvial Sedimentary Systems*, Proceedings of International Symposium (Alice Springs, Australia, September 2002), edited by F.J. Dyer, M.C. Thomas and J.M. Olley, IAHS Publication no. 276, pp. 211-218. Wallingford: IAHS Press.

- PAYRASTRE, O., GAUME, E. AND ANDRIEU, H., 2011. Usefulness of historical information for flood frequency analyses: developments based on a case study. *Water Resources Research*, 47 (8), W08511.
- PROSDOCIMI, I., KJELDSSEN, T.R. AND MILLER, J.D., 2015. Detection and attribution of urbanization effect on flood extremes using nonstationary flood frequency models. *Water Resources Research*, 51 (6), 4244-4262.
- PRUDHOMME, C. AND REED, D.W., 1998. Relationships between extreme daily precipitation and topography in a mountainous region – a case study in Scotland. *International Journal of Climatology*, 18 (13), 1439-1453.
- RANGER, N., MILLNER, A., DIETZ, S., FANKHAUSER, S., LOPEZ, A. AND RUTA, G., 2010. *Adaptation in the UK: a decision-making process*. Policy Brief for the Environment Agency, London.
- REED, D.W., 1987. Engaged on the ungauged: applications of the FSR rainfall-runoff method. In *Proceedings of the British Hydrological Society National Hydrological Symposium* (Hull, September 1987), 2.1-2.19.
- REED, D.W., 2002. Reinforcing flood risk estimation. *Philosophical Transactions of the Royal Society of London A*, 360 (1796), 1373-1387.
- REED, D.W., 2003. *Flood frequency estimation for Ouse at York: matched appraisal of flood events*. Report to Environment Agency Dales Area.
- REED, D.W., 2008. *Letcombe Brook: a hydrological review*. Report by DWRconsult for Faber Maunsell.
- REED, D.W., 2011. *Letters in Applied Hydrology*. DWRconsult, printed on demand at [www.lulu.com](http://www.lulu.com) [details at <http://www.lulu.com/shop/duncan-w-reed/letters-in-applied-hydrology/paperback/product-14681689.html>; accessed 18 October 2016].
- REED, D.W., 2014. *Independent review of Menston flooding problems*. Report to City of Bradford Metropolitan District Council, 11 December 2014.
- REIS, D.S. AND STEDINGER, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, 313 (1-2), 97-116.
- ROSBJERG, D., BLÖSCHL, G., BURN, D.H., CASTELLARIN, A., CROKE, B. DI BALDASSARRE, G., IACOBELLIS, V., KJELDSSEN, T., KUCZERA, G., MERZ, R., MONTANARI, A., MORRIS, D., OUARDA, T.B.M.J., REN, L., ROGGER, M., SALINAS, J.L., TOTH, E. AND VIGLIONE, A., 2013. Prediction of floods in ungauged basins. In *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales* (ed. G. Blöschl, M. Sivapalan, T. Wagener, A. Viglioni and H. Savenije), pp. 189-226. Cambridge: Cambridge University Press.
- ROYAL HASKONING, 2007. *Devon Flood Hydrology Strategy: report on additional analyses*. Report to Environment Agency South West Region.
- ROYAL HASKONING, 2013. *Devon Hydrology Strategy 2012 update*. Report 9X1663 to Environment Agency.
- SABOURIN, A. AND RENARD, B., 2015. Combining regional estimation and historical floods: a multivariate semiparametric peaks-over-threshold model with censored data. *Water Resources Research*, 51, 9646-9664.
- SEIBERT, J. AND BEVEN, K., 2009. Gauging the ungauged basin: how many discharge measurements are needed? *Hydrology and Earth System Sciences*, 13, 883-892.

- SLATER, L., 2016. To what extent have changes in channel capacity contributed to flood hazard trends in England and Wales? *Earth Surface Processes and Landforms*, 41 (8), 1115-1128.
- SOAR, P.J. AND THORNE, C.R., 2001. *Channel restoration design for meandering rivers*. ERDC/CHL CR-01-1. Washington DC: US Army Corps of Engineers.
- SOTO, A.U. AND MADRID-ARIS, M., 1994. Roughness coefficient in mountain rivers. In Proceedings of *Hydraulic Engineering 1994* (Buffalo, NY, 1–5 August 1994), edited by G. Cotroneo and R. Rumer. New York: American Society of Civil Engineering.
- STEDINGER, J.R. AND COHN, T.A., 1986. Flood frequency analysis with historical and paleoflood information. *Water Resources Research*, 22 (5), 785-793.
- STRUPCZEWSKI, W.G., KOCHANIEK, K. AND BOGDANOWICZ, E., 2014. Flood frequency analysis supported by the largest historical flood. *Natural Hazards and Earth System Science*, 14, 1543-1551.
- WALLINGFORD HYDROSOLUTIONS, 2016. *The Revitalised Flood Hydrograph model ReFH 2.2: technical guidance*. Wallingford: Wallingford HydroSolutions Ltd.
- WANG, Q.J., 1990a. Estimation of the GEV distribution from censored samples by method of partial probability weighted moments. *Journal of Hydrology*, 120 (1-4), 103-114.
- WANG, Q.J., 1990b. Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application. *Journal of Hydrology*, 120 (1-4), 115-124.
- WHARTON, G., 1992. Flood estimation from channel size: guidelines for using the channel-geometry method. *Applied Geography*, 12 (4), 339-359.
- WHARTON, G. AND TOMLINSON, J.J., 1999. Flood discharge estimation from river channel dimensions: results of applications in Java, Burundi, Ghana and Tanzania. *Hydrological Sciences Journal*, 44 (1), 97-111.
- WHARTON, G., ARNELL, N.W., GREGORY, K.J. AND GURNELL, A.M., 1989. River discharge estimated from river channel dimensions. *Journal of Hydrology*, 106 (3-4), 365-376.
- YEO, N. 2010. *A lichenometric and palaeohydraulic flood reconstruction of the East and West Lyn, Lynmouth*. Unpublished undergraduate dissertation, University of Southampton.



# List of abbreviations

|         |   |
|---------|---|
| AREA    | catchment drainage area (km <sup>2</sup> )                      |
| AMAX    | annual maximum  |
| BCW     | bankfull channel width  |
| BFI     | Base Flow Index   |
| BFIHOST | Base Flow Index estimated from soil type                        |
| CBHE    | Chronology of British Hydrological Events                       |
| CEH     | Centre for Ecology and Hydrology                                |
| CIRIA   | Construction Industry Research and Information Association      |
| CVRI    | coefficient of variation of recurrence intervals                |
| DPLBAR  | index describing catchment size and drainage path configuration |
| DPSBAR  | FEH index of mean drainage path slope                           |
| DTM     | Digital Terrain Model   |
| FARL    | FEH index of flood attenuation due to reservoirs and lakes      |
| FEH     | Flood Estimation Handbook                                       |
| FFA     | flood frequency analysis  |
| FPEXT   | FEH index describing floodplain extent                          |
| FSR     | Flood Studies Report  |
| GEV     | generalised extreme value distribution                          |
| GIS     | geographical information system                                 |
| GLO     | generalised logistic  |
| GW      | Geraldene Wharton   |
| HA      | hydrometric area  |
| HOST    | Hydrology of Soil Types   |
| IHDTM   | Integrated Hydrological Digital Terrain Model                   |
| IHM     | Integrated Height Model   |
| ISIS    | Hydrology and hydraulic modelling software                      |
| LiDAR   | light detection and ranging                                     |
| LMED    | median annual water level (with return period 2 years)          |
| MORECS  | Met Office Rainfall and Evaporation Calculation System          |
| MOSES   | Met Office Surface Exchange Scheme                              |
| NaFRA   | National Flood Risk Assessment                                  |
| NATMAP  | National Soil Map of England and Wales                          |

|            |  |
|------------|--|
| NERC       | Natural Environment Research Council             |
| NFCDD      | National Flood and Coastal Defence Database      |
| NRFA       | National River Flow Archive                      |
| OS         | Ordnance Survey                                  |
| PDM        | Probability Distributed Model                    |
| PE         | potential evaporation                            |
| POT        | peaks-over-threshold                             |
| PPWM       | partial probability weighted moments             |
| PROPWET    | FEH index of proportion of time that soil is wet |
| PWM        | probability weighted moments                     |
| QMED       | median annual flood (with return period 2 years) |
| RASP       | risk assessment for system planning              |
| ReFH       | Revitalised Flood Hydrograph method              |
| RHS        | River Habitat Survey                             |
| RMED       | median annual maximum rainfall depth             |
| RMSE       | root mean square error                           |
| RFFA       | regional flood frequency analysis                |
| SAAR       | standard average annual rainfall (mm)            |
| SMD        | soil moisture deficit                            |
| URBEXT     | FEH index of fractional urban extent             |
| VBA        | Visual Basic for Applications                    |
| WINFAP-FEH | Windows Frequency Analysis Package – FEH version |



# Appendix A: Palaeoflood techniques and data in UK river catchments

This independent review represents the views of the author, Mark G. Macklin,<sup>1,2</sup> and not necessarily those of the Environment Agency.

<sup>1</sup>River Basin Dynamics and Hydrology Research Group, Department of Geography and Earth Sciences, Aberystwyth University, Ceredigion, Wales, SY23 3DB, UK

<sup>2</sup>Innovative River Solutions, Institute of Agriculture and Environment, Massey University, Private Bag 11 222, Palmerston North, 4442, New Zealand

## A.1 Summary

This paper provides a review of palaeoflood techniques and data in UK river catchments for environmental management agencies and other water industry practitioners involved in flood risk assessment. In the context of rapidly emerging environmental impacts of anthropogenic climate change on extreme hydrological events, it concludes that traditional flood frequency analysis based on typically less than 50 year-long instrumental flow records in the UK need to be radically rethought in order to protect lives, property and nationally important energy and transport infrastructure assets. Event-based palaeoflood records are now available for all UK upland areas back to 1750 with more than 550 dated deposits associated with flood magnitudes approaching or exceeding those recorded in Boscastle in August 2004 and northwest England in December 2015. These records can be directly related to past and present short-term climate change and show a flood 'rich' period since 2007 associated with a shift in North Atlantic Oscillation phase most notably in the summer. In lowland floodplains event-scale palaeoflood records have recently been reconstructed for the last 4,000 years and also show that climate change is the main driver of flood frequency and magnitude. The reluctance of UK flood protection agencies to adopt and utilise palaeoflood studies arises from the misconception that these approaches are less precise and have a greater degree of uncertainty for assessing infrequent large floods than flood frequency analysis based on shorter term instrumental and documentary records. In practice, it is shown that there is as much if not greater uncertainty in assessing the frequency and magnitude of exceptional floods using these approaches. Whereas the palaeoflood record reconstructed from the fluvial sedimentary archive represents the only alternative for providing more data on the actual frequency and size of past floods over century and longer timescales. It is recommended all projections of future flooding in the UK, particularly those related to anthropogenic climate change, must incorporate palaeoflood data and approaches for risk assessment.

## A.2 Introduction

### A.2.1 Assumptions and limitations of traditional flood frequency analysis

Flood frequency analysis (FFA) for engineering design (Interagency Advisory Committee on Water Data, 1982) is based upon two assumptions: 1. 'annual maximum peak flows may be considered a sample of random and independent events' and, if a sufficiently long record is available, a frequency distribution for a site can be precisely determined; and 2. 'flood flows are not affected by climatic trends or cycles', which implies that climatic or environmental changes (e.g. catchment land cover or land-use) do not alter the statistical parameters of the frequency distribution – termed 'stationarity'. There is however growing realisation in the UK, and worldwide, that for assessment of flood risk associated with infrequent events of 1-0.01% annual probability these two basic assumptions of traditional FFA cannot be met. The first assumption – annual maximum peak flows are a sample of random and independent events – has been shown not to be true in the UK by growing evidence that both the frequency and magnitude of 1% and lower probability floods have changed significantly over time, particularly when the flood series is extended beyond the second half of the 20<sup>th</sup> century (Higgs, 1987; McEwen, 1987, 1990; Rumsby and Macklin, 1994; Longfield and Macklin, 1999; Macklin and Lewin, 2003; Macklin *et al.*, 2012 a,b; Macdonald, 2014). The second assumption of stationarity of flood flows also cannot be met because of hydroclimatic variability linked to shifts in atmospheric circulation (Rumsby and Macklin, 1994; Longfield and Macklin, 1999; Macklin and Rumsby, 2007; Foulds and Macklin, 2015), and that the second half of the 20<sup>th</sup> century (when most instrumental flows records started in the UK) was itself a period characterised by relatively small floods.

As a consequence of quasi-cyclic multi-decadal climatic fluctuations, notably the North Atlantic Oscillation (NAO; Walker, 1924) and the Atlantic Multi-decadal Oscillation (AMO; Schlesinger and Ramankutty, 1994; Kerr, 2000) arising from internal variability of the ocean-atmosphere system in the North Atlantic region, a single population of extreme flood events does not exist, nor is the probability of such extremes equal at any particular time. Anthropogenic climate change, induced by increasing green-house gas concentrations, further comprises the UK's civil engineering profession and regulatory agencies long-held notion that natural-world phenomena, including floods, fluctuate with a fixed envelope of statistical uncertainty that does not change over time (Boccaletti, 2015; Smith, 2015). Traditional FFA based on instrumental flow records of usually less than 50 years in length are therefore at best unlikely to provide robust estimates of flood events with a 1% or lower annual probability, and at worst are incorrect resulting in a significant under-estimate of flood risk. They cannot be extrapolated in order to predict the magnitude of an extreme event with a 0.01% annual probability, which is of particular concern to the Office for Nuclear Regulation (ONR) and the nuclear industry in the UK and worldwide (O'Connor *et al.*, 2014).

Neither statistical extrapolation nor stochastic re-sampling of monitored data reduces this uncertainty. Although there is still an active debate amongst hydrological statisticians in the civil engineering community (e.g. Serinaldi and Kilsby, 2015) of the merits or otherwise of nonstationary models in hydrologic frequency analysis, these studies focus solely on the evaluation of instrumental records over time periods rarely exceeding 100 years. Extending the observational record by using historical and geological archives is the only alternative for providing more data on the actual frequency and size of past floods.

## A.2.2 Enhancing flood frequency analysis using documentary and palaeoflood records

The use of documentary evidence of historical flood events in contemporary flood frequency estimation has significantly expanded in the UK over the last 25 years (Archer, 1987; McEwen, 1987; Rumsby, 1991; Rumsby and Macklin, 1994; Longfield, 1998; Longfield and Macklin, 1999; Macdonald, 2006). Documentary records most often predate the installation of gauging stations and usually provide indirect information on peak flood discharge, often in the form of a water-level marker, information that a specific location had been flooded, damaged or destroyed, or that a flood reached a level relative to a structure (Kjeldsen *et al.*, 2014). There are, however, three well recognised limitations of documentary records in the UK, in terms of flood frequency and particularly flood magnitude analysis (Rumsby, 1991). First, the accuracy and reliability of measurements deteriorates before AD 1700. Second, the record is biased towards populated areas. Third, and most significant, changes in channel capacity resulting from river aggradation and incision (Macklin *et al.*, 2010, 2013), floodplain morphology through sedimentation and wetland drainage (Lewin and Macklin, 2010; Macklin *et al.*, 2010, 2014) and construction of bridges, embankments and transport infrastructure (Lewin, 2010, 2013) make it very difficult to convert a historical water level to a peak flood discharge. These changes, in conjunction with floodplain encroachment that accelerated in the nineteenth century, have affected to greater or lesser extent waterways in all major UK towns and cities since the Industrial Revolution (Lewin, 2013) with major channel alterations, in some areas, beginning as early as the medieval period (Lewin, 2010).

Although peak discharges of major floods have been reconstructed using historical water level information (e.g. Lewes - Macdonald *et al.*, 2014 - and York - Macdonald and Black, 2010 - back to AD 1750 and AD 1263, respectively), as consequence of major channel and floodplain modifications they are likely to have large, unsystematic and presently unknown errors. In the case of York, historical maps and documentary accounts back to c. AD 1750 are the only basis for constraining channel position and these give no information on channel capacity, particularly river bed-level fluctuations, or changes in floodplain elevation and morphology. As shown by archaeological excavations and sedimentological investigations undertaken in York during the construction of the North Street pumping station during 1993 (Hudson-Edwards *et al.*, 1999), the Yorkshire Ouse was affected by more than 9 m of channel-margin sedimentation between Roman times and the early modern period, which makes the assumption of constancy in channel capacity since the 13<sup>th</sup> century AD when documentary flood reconstructions at York begins, likely to be erroneous. Similarly in Lewes, although flood series extension for the Sussex Ouse goes back only to 1750 and periods of bridge construction and rebuilding are known, no information is provided on channel cross-section changes over the last 250 years. Indeed, the authors state that 'inevitably the potential for modification to the channel cross section during the historical period represents a challenge when estimating historical flows' (p. 2819, Macdonald *et al.*, 2014). Moreover, while the position of the Sussex Ouse channel at Lewes appears to changed little since the first Ordnance Survey of 1875, these and later maps show major changes on Ouse floodplain as a result of construction and re-development.

These fundamental limitations of documentary based flood-level records concerning channel cross section and floodplain elevation/morphology changes that are especially evident in urban contexts, could be addressed if they were assessed in conjunction with geomorphological investigations at these sites, which can provide information on channel and floodplain evolution. By contrast, the majority of UK palaeoflood studies in both upland (e.g. Macklin *et al.*, 1992; Macklin and Rumsby, 2007; Merrett, 2001) and lowland (e.g. Jones *et al.*, 2010, 2012) have also documented short- and long-term

changes in vertical tendency of river channels, variations in channel form and capacity, as well as periods of floodplain sedimentation.

Unfortunately in the UK - with some notable exceptions in northern England (Rumsby, 1991; Merrett, 2001; Foulds, 2008), Wales (Jones, 2007) and Scotland (Werritty *et al.*, 2006) - instrumental, documentary and palaeoflood records have not been investigated in an integrated manner. This is in strong contrast to the USA (e.g. O'Connor *et al.*, 2014) and mainland Europe (e.g. Benito *et al.*, 2008; Toonen *et al.*, 2015) where combined studies of instrumental, documentary and geologic records of major floods are becoming increasing routine and are being used by regulatory and environmental protection agencies to inform flood risk assessment. The UK lags significantly behind this rapidly developing field of water-resource risk assessment and planning.

### **A.2.3 Purpose and structure of review**

The principal aim of this paper is to critically review palaeoflood studies in the UK and, in the context of the rapidly emerging environmental impacts of anthropogenic climate change on extreme hydrological events (Lewin and Macklin, 2010), recommend that they need to become part of routine flood risk assessment. This becomes a matter of both importance and urgency in the light of the major transport (e.g. High Speed 2 and 3 rail networks; Heathrow Airport's proposed third runway), urban (e.g. Northern Powerhouse) and energy (e.g. new nuclear power plants, including Hinkley Point C) infrastructure developments that are currently being planned and will shape UK society in the 21<sup>st</sup> century. There are no constraints from a practical, methodological or cost viewpoint why palaeoflood hydrology using geomorphological and sedimentary archives should not be included in these large-scale and other floodplain developments where property and lives are at risk. The reluctance of the UK civil engineering and hydrology academics, practitioners, and water regulatory agencies, to adopt and use palaeoflood techniques and data seems to be partly a disciplinary 'silo' effect but also arises from the wish or necessity of having to use using 'the industry standard' in flood risk assessment. The latter approach is becoming increasingly untenable when so many recent extreme floods are described as 'unprecedented' from the perspective of the less 50 year-long instrumental record even though palaeoflood records in same catchment show that much larger events occurred in the last 100-200 years (Foulds *et al.*, 2013, 2014; Foulds and Macklin, 2015). This view has been further underscored by the December 2015 Storm Desmond floods in northwest England, and it could now be contended that the water management and risk assessment agencies in the UK are not presently exercising due diligence by failing to incorporate palaeoflood techniques and data in FFA.

This paper is divided into three major sections. In the first part event-scale palaeoflood records and flood recording sedimentary environments in the UK are reviewed. In the second section the UK Environment Agency's current approach to FFA in the context of climate change is critically assessed, as it is likely to affect major energy, transport and urban infrastructure development. Finally, some anticipated 'frequently asked questions' from UK water practitioners, regulators and risk managers concerning palaeoflood techniques and data are addressed.

## A.3 Geologically extended histories of UK riverine floods using palaeofloods

### A.3.1 Introduction

Geologic records can extend the knowledge of rare hazards from floods, storm surges and tsunamis (O'Connor *et al.*, 2014). In riverine environments these types of investigations have been termed palaeoflood hydrology (Kochel and Baker, 1982) and are defined as 'the reconstruction of the magnitude and frequency of past floods using geologic evidence' (Baker *et al.*, 2002). Palaeoflood studies were pioneered in the USA where they are now used routinely to provide reliable estimates of rare floods (potentially with annual exceedence probabilities of  $10^{-6}$ ) for critical structures such as dam spillways, nuclear power-plants and hazard waste repositories (O'Connor *et al.*, 2014). Layered sedimentary deposits commonly give the most complete geologic record of large floods, and they may be preserved for hundreds or thousands of years in suitable environments, thereby providing an archive of rare, high-magnitude events (Benito and O'Connor, 2013).

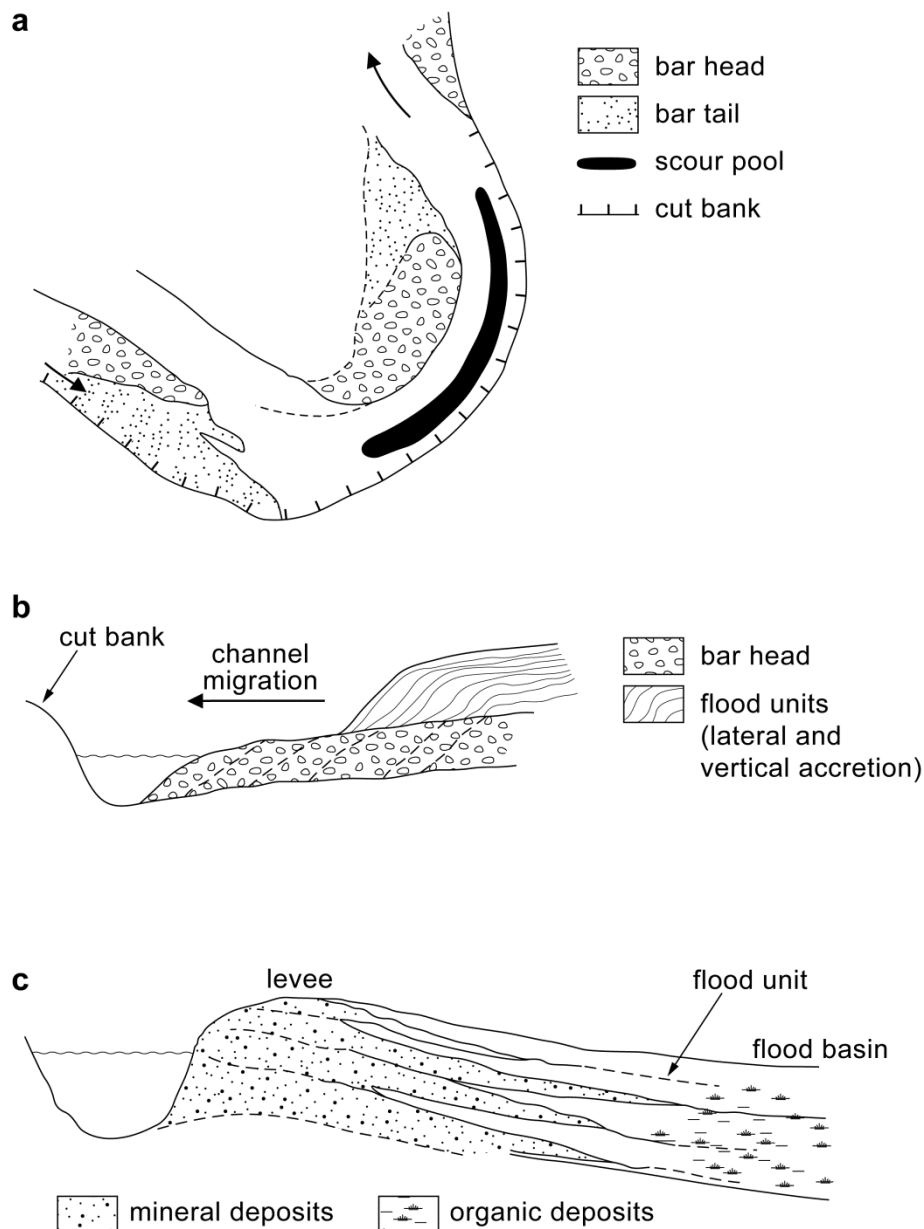
Despite UK scientists leading research into the geological archives of floodplains for quantifying the impacts of historical (Macklin *et al.*, 1992a, b; Carling and Grodek, 1994; Rumsby and Macklin, 1994; Merrett, 2001; Macklin and Rumsby, 2007) and longer-term (Macklin *et al.*, 2012b) climate change, palaeoflood studies have as yet not been used by engineers and flood protection agencies for FFA in the UK. The primary reasons appear to arise from the misconception that palaeoflood approaches are less precise and have a greater degree of uncertainty for assessing infrequent large floods than FFA based on short-term instrumental records, but also from the over-reliance on model-based studies that (incorrectly) assume stationarity in the flood series.

This paper examines the Holocene flood series sedimentary record for the UK, and considers whether types of evidence and analysis that as yet remain little exploited could be of value - in the specific and distinctive (though not unique) context of UK alluvial environments. This is a more focused aim than would be involved in summarizing Holocene alluvial stratigraphy, or even of palaeohydrological interpretation of sediment units. The approach also differs from the meta-analysis of statistically significant peaks in Holocene  $^{14}\text{C}$ -dated sedimentological changes in fluvial environments that represent centennial-scale flood episodes (Macklin and Lewin, 2003; Macklin *et al.*, 2005, 2010, 2012a; Benito *et al.*, 2015). Here only the more limited record of *event sequences* is considered at specific sites whose flood series are chronologically well constrained.

### A.3.2 Flood sediments

Alluvial deposits of rivers represent an unwritten flood record. The floodplain sediments that get preserved are deposited almost entirely during and following successions of extreme events, both by overbank flows out across floodplains and by channel-zone deposition of coarser bed materials and of finer sediment in slack-water environments. Floods rise and fall, and characteristically leave behind a sediment signature. This is generally heterolithic in character, which is they incorporate sequences or couplets both of coarser material from peak discharges and finer material from waning flows or inter-flood discharges. In channel zones, flood mobilization of coarse materials may lead to both channel erosion and deposition producing meso-scale bed and bar forms (Figure A.1a). Second, where sediment sheets are deposited at channel margins at various flood stages, the generally finer suspended-sediment units may exhibit textural changes, usually in the form of smaller scale coarse to fine grading marking the change from flood peak to waning stage. Unit thickness may relate to flood duration and

magnitude, but also to intra-flood sediment loadings and to flood sequence sediment-exhaustion effects.



**Figure A.1 Flood sediment units showing (a) in-channel bedforms, (b) channel margin accretion and (c) floodplain depositional sites**

Source: Jones et al. (2010b)

Units tend to drape across existing topography so that they may be inclined obliquely at channel margins (on a so-called 'inner accretionary bank' on the inside of meander bends on top of previously deposited bedforms, for example) or horizontally on floodplain surfaces, or confined to post-flood ponding sites (Figure A.1b). Within channels, lesser flows may leave bands of finer material between the coarser sediment incursions from larger floods. With iteration, this may in time lead to thick but visually undifferentiated fine-sediment units, possibly with only the most extreme events leaving a clear individual sediment signature. Finally, out on floodplains, material may occasionally disperse to form a surface drape or incursion of renewed deposition into otherwise stable, soil-forming domains or into organic floodbasin depositional environments (Figure A.1c). Some individual flood deposits are of substantial thickness

and easily recognised, whereas others may be little more than a surface veneer without heterolithic differentiation and which dries out and becomes rapidly obscured.

There is also a spectrum of sedimentary environments present in channels and on floodplains and these remain openly receptive for flood deposition to differing extents (Lewin *et al.*, 2005). Single sections through flood deposits are likely to record flooding for variable and limited time periods, some for a single event, and others for many decades. As in dendrochronology, a more extended record needs to emerge by matching and collating sets of discrete or (preferably) overlapping records. Furthermore, in-channel sediments may relate to sub-bankfull flows, channel dimensions themselves to channel-forming discharges (commonly, but not in all environments accurately, ascribed to an average recurrence interval of 1-2 years), whilst floodbasin flood units may be generated by rarer extreme events that are the only ones to transport sediments so far from the river.

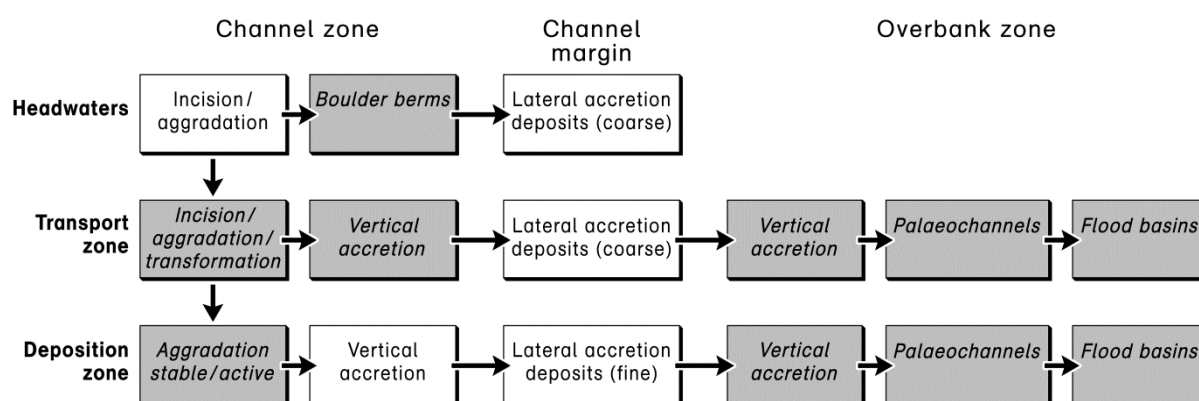
For flood record purposes, it is helpful if palaeoflood indicators are diagnostic of a range of flood magnitudes and frequencies. The presence of flow-magnitude dependent sediment sizes is important. For bedload-transport materials at some locations, this allows (coarser) extreme-event materials, whose size is related to incident stream-power and is competence-limited, to establish the range of flows that transported them (Maizels and Aitken, 1991). For the recording of out-of-channel extreme events and finer sediment deposition this is not the case. The relative absence of sandy sediments in many UK rivers is also something of a disadvantage, because most overbank deposited materials are derived from eroded soils and these get widely deposited from turbid waters whatever the magnitude of flood involved. Some geological materials in Britain yield a diagnostic range of sediment sizes (e.g. catchments underlain by mixed Carboniferous rock types – Macklin *et al.*, 1992b), but unfortunately for flood-power recording, many do not.

### **A.3.3 Flood recording sedimentary environments in UK river catchments**

Table A.1 lists flood recording riverine sedimentary environments which have been or might be used to provide data for a UK Holocene flood history. They are variably available within river catchments, and Figure A.2 illustrates these in terms both of local depositional environment (channel, channel margin and overbank sites) and catchment location (headwater, transport and depositional zones). For datable units recording single floods, the most useful so far have proved to be boulder bars and berms in headwater environments (Macklin *et al.*, 1992a; Merrett and Macklin, 1999; Macklin and Rumsby, 2007; Foulds *et al.*, 2013, 2014, 2015), in-channel vertical accretion deposits formed where enlarged channels have contracted (Macklin *et al.*, 1992b; Rumsby, 2000), palaeochannel fills (Werritty *et al.*, 2006; Jones *et al.*, 2011, 2012), and floodbasin incursions (Tipping, 1998; Jones *et al.*, 2010a, 2012). For stable lowland river environments, the range of flood-recording types of unit is rather less, and concentrated on heterolithic overbank sequences.

**Table A.1 Riverine sedimentary environments in the UK where event-scale palaeoflood records have been reconstructed**

| Sedimentary environment           | References   |
|-----------------------------------|--|
| <b>Channel and channel margin</b> |  |
| Vertical accretion units          | Macklin et al. 1992b, Rumsby 2000  |
| Boulder berms and bars            | Macklin et al. 1992a, Merrett and Macklin 1999, Macklin and Rumsby 2007, Foulds et al. 2014a, Foulds et al. 2014b, Foulds and Macklin 2016 |
| Lateral accretion units           | Brown et al. 2001, Brown 2008  |
| <b>Overbank</b>                   |  |
| Palaeochannel fills               | Werritty et al. 2006, Jones et al. 2011, Jones et al. 2012   |
| Floodbasin incursions             | Tipping 1998, Jones et al. 2010  |



**Figure A.2 Flood recording environments**

Notes: Those used in the UK are shaded.

Source: Jones et al. (2010b)

Lake sediments have also recently been evaluated as a potential source of palaeoflood data in the UK (see review by Schillereff *et al.*, 2014) but presently at only one site – Brotherswater (northwest England, catchment area 13 km<sup>2</sup>) – has palaeoflood histories been reconstructed with radiometric dating control back to 1960 (Schillereff *et al.*, 2015). But because of uncertainties in analytical resolution and the age-depth model, multiple floods that occurred during the broadly annual timespan of individual samples cannot be separated in the lake stratigraphic record. They conclude that establishing flood magnitudes from lake sediments remains a great challenge, which until addressed will preclude their use in FFA.

A yet unpublished (Schillereff, 2015) record of flooding in the Brotherswater catchment, which is believed to extend back to the 7<sup>th</sup> century AD, is of particular interest and can be used to evaluate the utility of lakes in upland UK to extend the flood series beyond instrumental and documentary records. Unfortunately, because of significant problems with the <sup>14</sup>C dating of their cores (6 of their 13 <sup>14</sup>C dates are too old as a result of eroded soils delivering ‘old carbon’ into the lake) a robust flood chronology cannot a



present be extended beyond the 1860 and even during the 19<sup>th</sup> century the catchment was affected by mining activities. Major age reversals in the cores shortly after AD 1350 indicate that lake sedimentation rates have significantly influenced by landuse change and farming, which makes palaeoflood reconstruction with event scale resolution at this site problematic. Although similar to meta-analysis of Holocene <sup>14</sup>C-dated flood units in riverine contexts (e.g. Macklin *et al.*, 2012a) lake records may be useful for identifying centennial length and longer flood 'poor' and flood 'rich' periods.

Given the very limited number of lake-based palaeoflood studies in the UK, together with dating uncertainties, problems with event resolution and human impacts of sedimentation rates, they do not currently constitute, as claimed, a more complete flood record compared to the many hundreds of securely dated flood events that are available from riverine sedimentary contexts. Furthermore, a recent analysis of the UK Holocene <sup>14</sup>C dated fluvial record has established when and where human activities effected hydrological and sedimentation regimes of river catchments (Macklin *et al.*, 2010, 2014). AD 1050-1350 emerges as the period during which human impact on fine sediment supply to rivers (measured by the formation of physically distinctive anthropogenic sediments - anthropogenic alluvium) was at its greatest during the Holocene, both for the UK as a whole and also in northwest England (Macklin *et al.*, 2014). This again points to the flood 'rich' phase identified by Schillereff (2015) in Brotherswater at AD 1100-1300 to have been affected by increased sediment supply as a consequence of catchment landuse change. But as first recognised by Macklin (1999), deforestation and agricultural practices by increasing runoff and sediment supply have rendered river catchments in the UK to be more sensitive to climate change. Also higher channel margin vertical accretion rates have increased the preservation of flood units, especially those associated with large floods (Macklin *et al.*, 1992a; Rumsby, 2000).

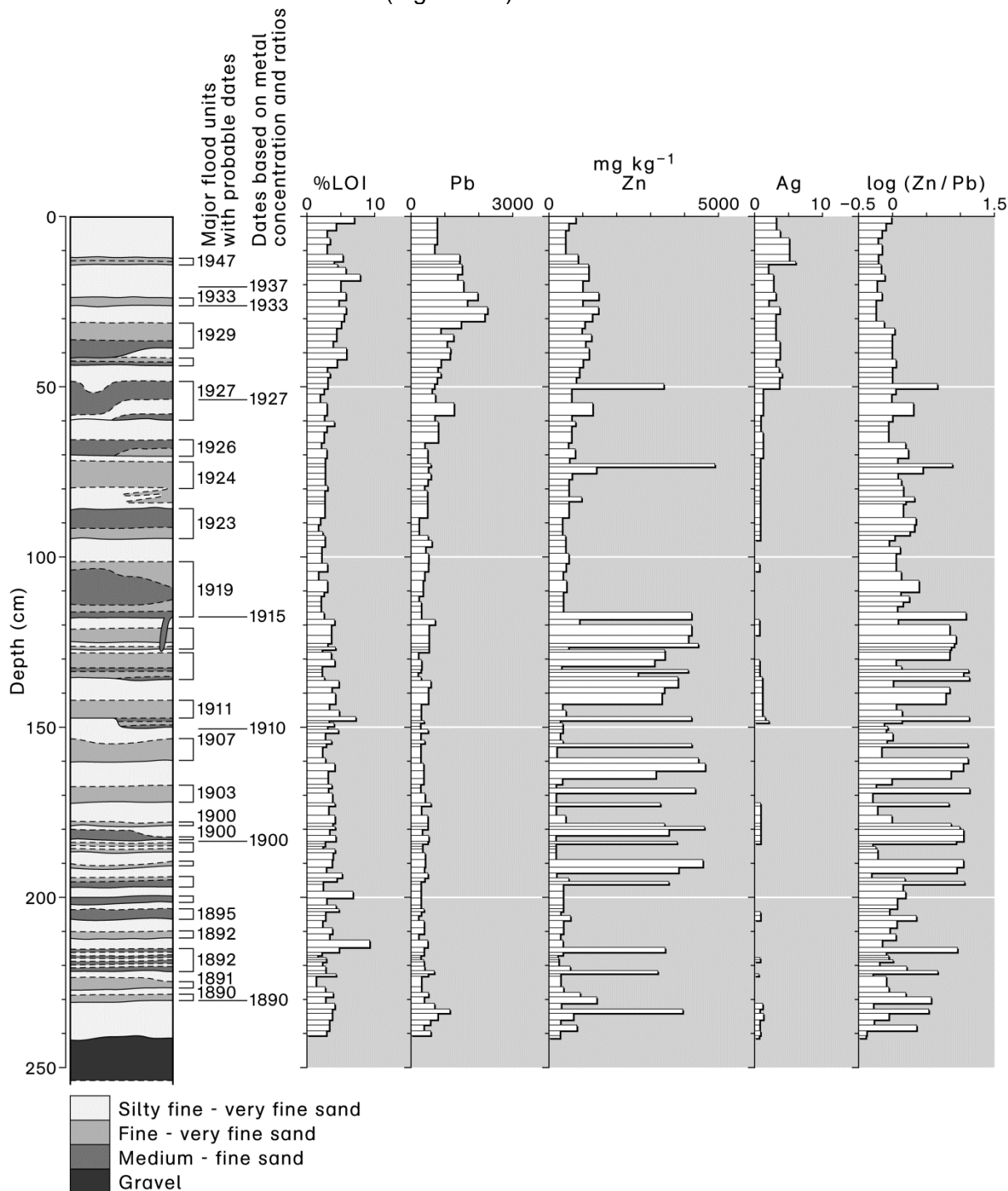
## *Channel and channel margin*

### Boulder berms and bars

In upland catchments of northern and western UK very large and powerful flows of water, similar to that affected Boscastle in 2004, deposit boulders and large cobbles on floodplains. Termed 'boulder berms' (Macklin *et al.*, 1992a), these can be dated using lichenometry, which is based on the assumption that when a boulder comes to rest after a flood, lichens are able to colonise newly exposed rock surfaces within a few years. Their growth rates, based on rock surfaces of known age – such as gravestones and built structures – can be used to calculate an age for a flood event (Macklin and Rumsby, 2007). This method has been used to reconstruct records of extreme flood events in the Northern Pennines (Rumsby, 1991; Macklin *et al.*, 1992a), Lake District (Johnson and Warburton, 2002), Yorkshire Dales (Merrett and Macklin, 1999; Merrett, 2001) Cambrian Mountains (Foulds *et al.*, 2014), Brecon Beacons (Macklin and Rumsby, 2007) and Dartmoor (Foulds *et al.*, 2013). Comprising 556 dated deposits extending back to 1750 in most regions (the length of record is primarily limited by lichen longevity, which is approximately 350 years in upland UK) and even to the 11<sup>th</sup> AD on some areas where boulder berm age is constrained by <sup>14</sup>C dating (Macklin *et al.*, 2013), this now provides one of the longest and complete databases of upland flooding in Europe (Foulds and Macklin, 2015). Boulder size measurements and hydraulic modelling techniques can also be used to estimate discharges (Carling and Grodek, 1994; see Merrett, 2001 for a detailed review), so in conjunction with lichen-based dating, event-based flood magnitude-frequency relationships can be reconstructed over multi-centennial length timescales.

## Vertical accretion units

In the lowland valley of the River Tyne, vertically-accreting flood sediment units in a contracting channel context have been studied in some detail (Macklin *et al.*, 1992b; Rumsby, 2000). These formed heterolithic sequences of alternating sand and silty-sand units dated to AD 1890-1947 (Figure A.3).



**Figure A.3 Flood units at Low Prudhoe, River Tyne**

Source: Macklin *et al.* (1992b)

The sedimentary record was dated using contaminant lead and zinc concentrations resulting from historical metal mining in the Tyne catchment (Macklin *et al.*, 1994) and linked to documented flood events (Macklin *et al.*, 1992b). This gave a set of 25 large floods, best recorded in the earlier part of the period before sediment-build up reduced deposition from any but the largest events. Further sections from Broomhaugh Island

on the Tyne were presented by Rumsby (2000) extending the flood sequence back to AD 1600 and forward to AD 1939. Self-censoring is a feature of many flood-sediment depositional environments as accommodation space becomes filled or inaccessible so that channel contraction provides a temporally limited record. However, contraction episodes may have occurred other than in the recent past, as was suggested by Brown *et al.* (1994) for the earlier Holocene, so other sites are likely to exist but are as yet unrecorded.

## Lateral accretion units

Holocene gravels resulting from lateral accretion over a longer period, and which incorporate dated and flood-destroyed bridge structures and other archaeological materials, have been studied in considerable detail on the River Trent at Hemington (Brown *et al.*, 2001). Bridge destruction episodes come within an extended period of medieval and later flood documentation and of channel instability, although it is difficult to date the gravels themselves despite the quantity of dated organic and mineral materials (Brown, 2008). Reconstruction of bridge-destroying flows (in Brown, 2008) suggested magnitudes not dissimilar to present-day bankfull flows on the Trent and an ongoing history of flood-related bed scour and lateral channel change, though not one that as yet directly provides an independent flood-event and magnitude sequence.

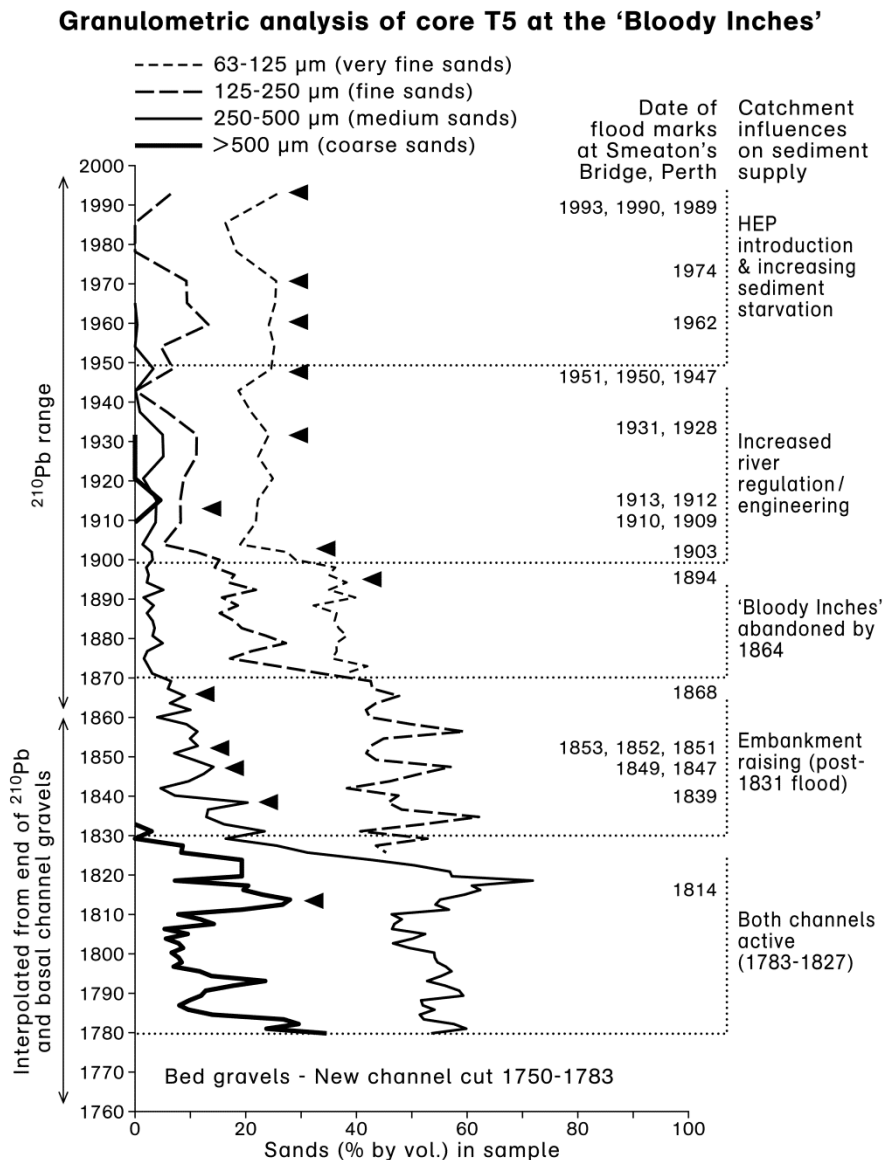
## Overbank

## Palaeochannel fills

Werritty *et al.* (2006) have similarly examined flood sequences recorded in alluvial cutoff sediments. A loop of the River Tay began to be abandoned in around 1761, and radiometric dating of the cored sediment fill allowed documented floods and the heterolithic sequence to be related over a period of around 200 years (Figure A.4). Coarser sediment units related well to known nineteenth century floods, with matching becoming increasingly difficult as silts replaced sands in the upper part of the fill. For any individual site, recording potential is again of limited duration because the sedimenting void becomes filled, so that for a more extended history, sets of overlapping flood-sedimenting cutoff sites are needed. But many  $^{14}\text{C}$ -dated palaeochannels do exist: Macklin *et al.* (2010) record 327 palaeochannel sites in the UK out of a total of 776 dated Holocene alluvial units, a number that continues to increase with further study (Macklin *et al.*, 2012a). So far those dated are skewed towards more recent millennia. This may reflect better preservation (earlier fills having been removed by erosion), more active cutoff formation in recent times, and at least to some extent the present state of discovery.

Toonen and his co-workers (Toonen *et al.*, 2015) have recently reconstructed event-scale flood peaks of the Lower Rhine for the last 450 years from grain-size measurements of flood deposits within abandon channels and dyke breach scour holes. Grain size descriptive measures such as the 95<sup>th</sup> percentile and end-member modelling correlate well with instrumental flood peaks and were found to provide sensitive proxies for flood magnitudes. In a related study also in the Lower Rhine, Toonen *et al.* (2013) using a slope-area approach and Chézy-based hydraulic model estimated the magnitude of extreme floods dating to between 4100 and 7900 years before present from clay flood units within a peaty channel fill of an early Holocene – Lateglacial terrace. The modelled minimum discharges for these events is  $13,250 \text{ m}^3 \text{ s}^{-1}$  that exceed all gauged records and have a minimum recurrence of between 1,250 and 2,500 years.

Investigations of this type have not yet been undertaken in the UK but both historical and Holocene age palaeochannels on terraced floodplains are well developed in all regions and major catchments of the UK (Macklin *et al.*, 2010, 2013). Preliminary studies in the Afon Tywi (Jones *et al.*, 2011) has shown that there is enormous potential to develop and apply palaeoflood approaches based on palaeochannel slackwater deposits in the UK.



**Figure A.4 Cut-off fill flood units from Bloody Inches, River Tay**

Source: Werritty *et al.* (2006)

## Floodbasins

Floodbasins, located behind natural levees, are the second type of depositional environment where palaeoflood studies can be undertaken in lowland UK floodplains (Jones *et al.*, 2010a, 2012). Floodplains, including those that are influenced by tidal flows, and adjoining river terraces typically have the longest records of slackwater deposits left by large floods at sites of persistent sediment accumulation. Such records can extend back several thousand years in protected environments. Large floods in lowland floodplains can be identified by coarser sedimentary layers that reverse fining-

upward sequences in floodplain deposits (Macklin *et al.*, 1992b). These flood units can be dated directly when comprised of coarse silt and sand using Optically Stimulated Luminescence (OSL) dating techniques (Duller, 2004), or bracketed by radiocarbon dating of organic material incorporated within floodplain sediments (Macklin *et al.*, 2010).

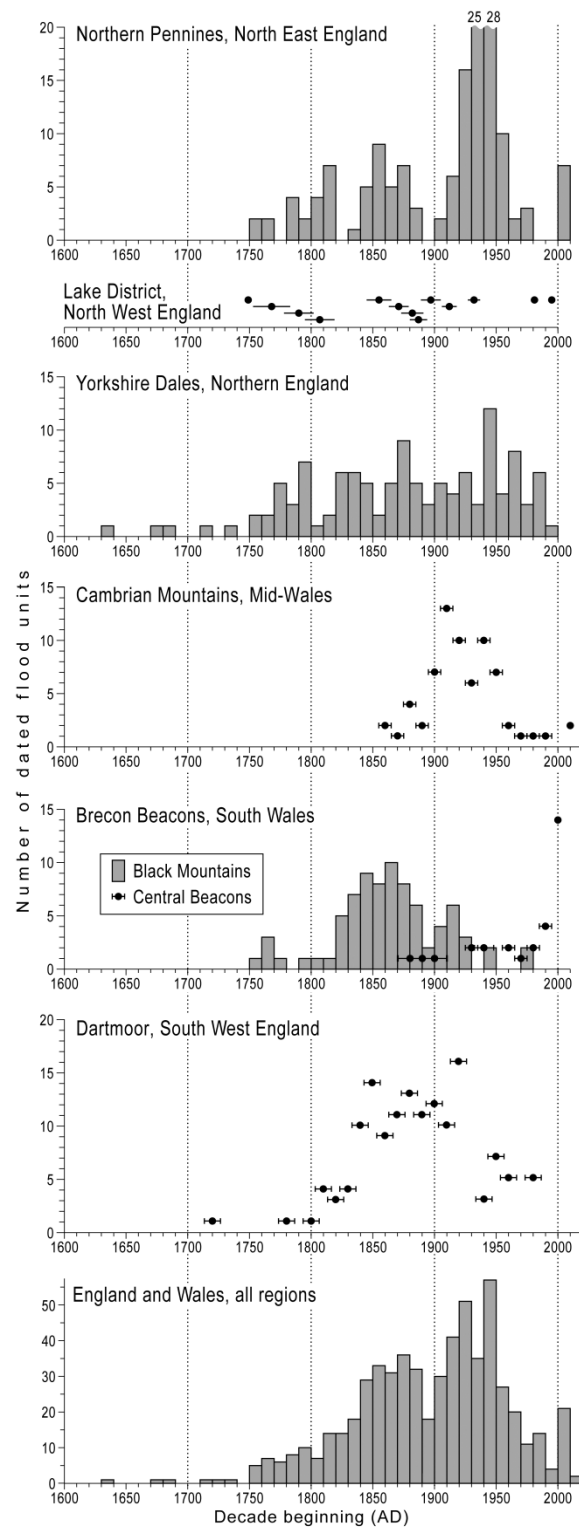
Flood basin deposits do appear to cover a more extended earlier Holocene timespan. Such sediments were cored and dated by Tipping (1998) on the River Glen in Northumberland, demonstrating a limited number of mineral incursions into peaty deposits. Jones *et al.* (2010a, 2012) similarly dated some 11 floodbasin flood units on Welsh rivers. These correlated with independently dated flooding episodes previously identified from  $^{14}\text{C}$ -dated sedimentation changes in a totalled set of UK Holocene alluvial units (Macklin *et al.*, 2010, 2012a). Floodbasin deposits represent a censored flood record, because only the very largest floods are sufficient to deposit identifiable (coarser) sediment units.

### A.3.4 Event-scale palaeoflood records

Studies of the UK boulder berm sedimentary archive (Foulds and Macklin, 2015) show that 21<sup>st</sup> century floods are not unprecedented either in terms of their frequency (large floods were more common before 1960; Figure A.5) or magnitude (the largest flood events occurred during the 17<sup>th</sup> – 19<sup>th</sup> centuries; Figure A.6). However, in some areas (Northern Pennines, Brecon Beacons) recent floods have equalled the largest historical events. The occurrence of extreme upland floods, and indeed all large floods in the UK, is strongly related to the phase of the NAO (Macklin and Rumsby, 2007; Foulds *et al.*, 2013, 2014; Foulds and Macklin, 2015) with winter rain-on-snow and torrential summer downpours associated with negative NAO index values. Rapid warming of the Arctic in recent years relative to the mid-latitudes – so called Arctic amplification – may also be responsible for creating a ‘wavier’ jet stream (Francis and Vavrus, 2015), leading to more persistent weather patterns associated with extreme hydrological events.

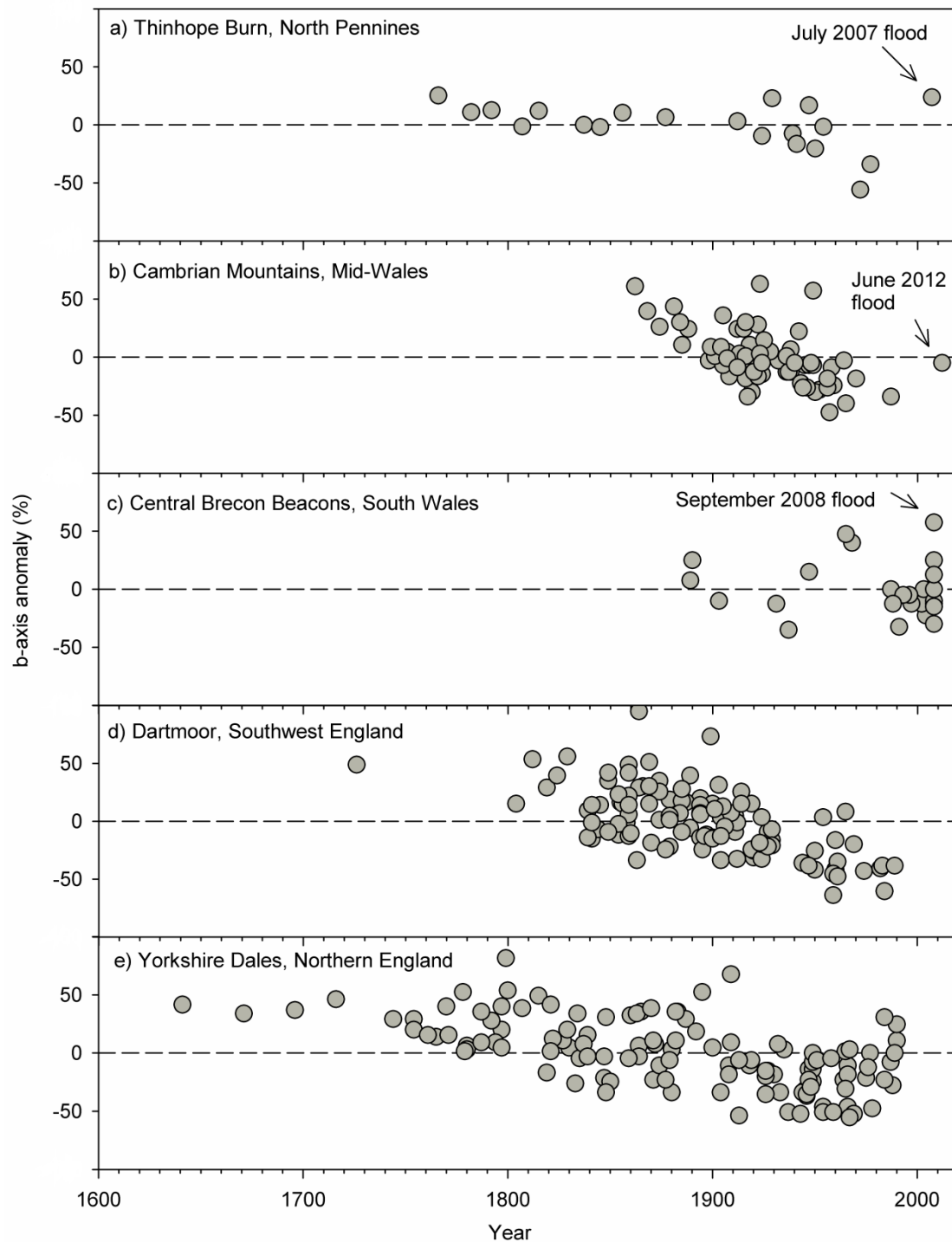
Hydraulic modelling techniques and lichen dating of boulder berms and bars can provide reliable multi-centennial length reconstructions of extreme flood peaks (Rumsby, 1991; Merrett, 2001). In Coverdale, North Yorkshire, for example the peak flow discharge of 26 events was reconstructed between 1670 and 2000 (Merrett and Macklin, 1999; Merrett, 2001; Macklin and Lewin, 2008). Similar detailed reconstructions of flood magnitude back to the 18<sup>th</sup> century have been undertaken in more than 30 catchments in the Northern Pennines (Rumsby, 1991) and the Yorkshire Dales (Merrett, 2001). With the more recent development of airborne and ground-based LiDAR, as well as terrestrial laser scanning and structure from motion imaging techniques, high resolution digital elevation models (DEMs) can now be constructed. These can facilitate high resolution modelling of extreme flood events using boulder berm and bar deposits in most upland catchments in the UK.

In the upper Severn catchment, mid-Wales, a record of overbank flooding has been reconstructed for the last 3750 years (Jones *et al.*, 2012; Figure A.7). This shows that multi-centennial length periods characterised by the more frequent occurrence of high magnitude floods have alternated with periods of similar length without such floods. These periods correspond to large-scale fluctuations in hydroclimate, most notably multi-centennial and multi-decadal variations in NAO. Between AD 1100 and 1300, a time of warmer temperatures and more positive NAO, large floods of  $c. \leq 3\%$  annual probability were very rare with only two events recorded in  $c. 250$  years. By contrast before AD 1000 and particularly after AD 1550, during the cooler Little Ice Age, the frequency of large floods increased significantly. The largest floods during the past 3750 years occurred at  $c. 235$  and 10 BC, and exceeded the flood of 12<sup>th</sup> February 1795 that was the largest since 1672 when flood levels were first recorded in the River Severn.



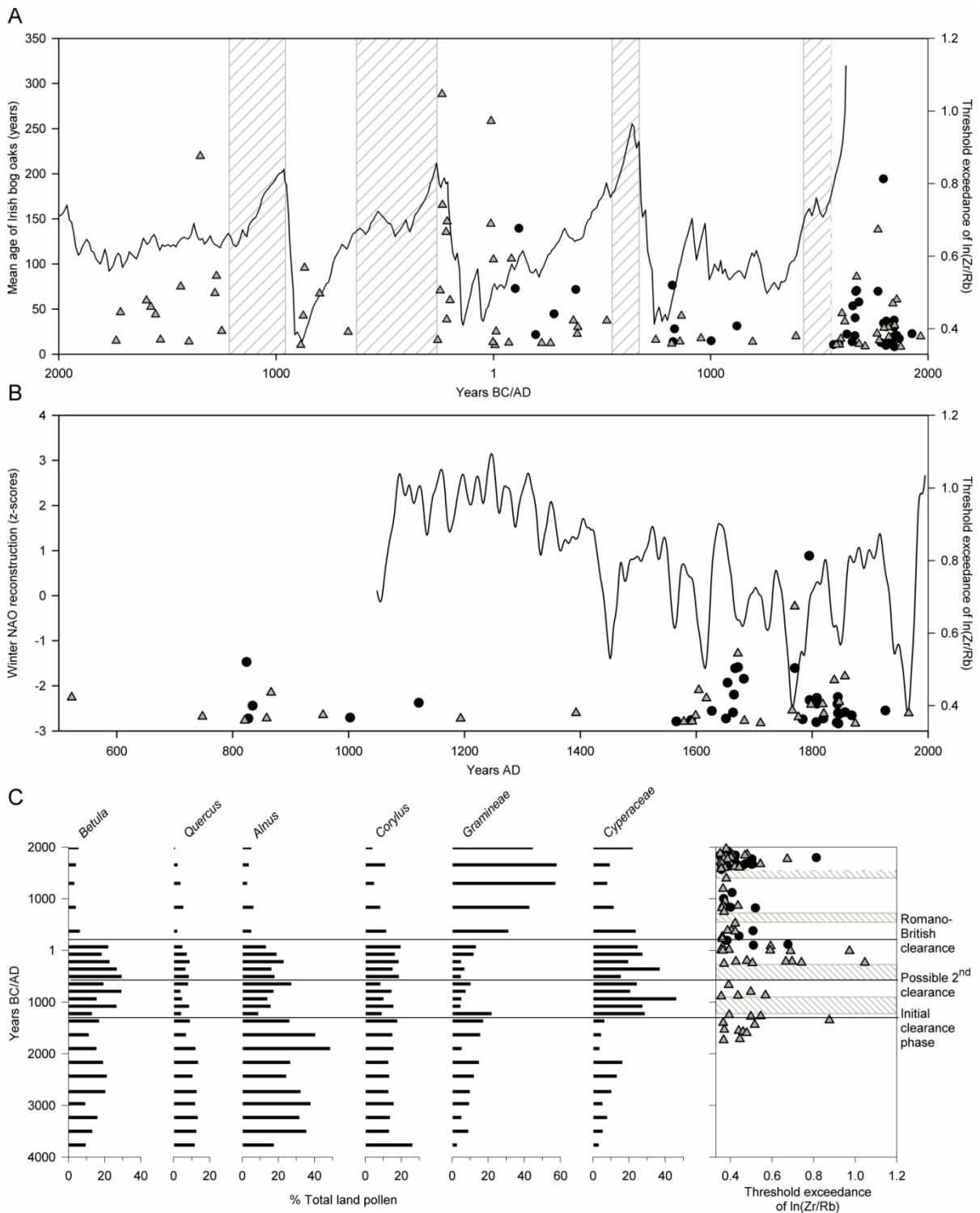
**Figure A.5 Decadal frequency of lichen-dated boulder berms in upland areas of England and Wales**

Source: Foulds and Macklin (2016)



**Figure A.6 Relative flood magnitude based on average boulder berm b-axis measurements**

Notes: '0' on the y axis (dashed line) represents the average size of boulders moved by extreme floods in each study catchment.  
Source: Foulds and Macklin (2016)



**Figure A.7** 3,700-year record of major floods (events plotted equate to  $\leq 3\%$  of present annual exceedance probability) in the upper Severn Wales: (A) comparison with mean age of Irish bog oaks (Leuschner et al. 2002), a regional hydroclimate proxy; (B) comparison with reconstructed winter NAO index (Trouet et al. 2009); and (C) comparison with the pollen record from Carneddau in the headwaters of the Rhiw catchment (Walker 1993)

**Notes:** In the more recent part of the record, the same flood may be represented in the data from core 1 (grey triangles) and from core 2 (black circles). In (A), grey shading highlights periods characterised by an absence of major floods and by either a peak or a sustained increased in the mean age of Irish bog



oaks. Note the correspondence between steady increases in the mean age of Irish bog oaks and an absence of major floods.

Six taxa are shown in (C) from which the author has identified 2 episodes of woodland clearance and a possible third episode between the other 2. Grey shading is as in (A).

Source: Jones *et al.* (2012)

One of the most significant recent findings from the studies of the floodplain sedimentary archive is that some of the very largest floods over the last 5000-10,000 years – termed millennial floods – can be considered to unique events ('perfect storms') and do not necessarily relate to climate change (Macklin *et al.*, 2013; Toonen *et al.*, in press). This underscores the importance of bespoke, site based palaeoflood studies at all existing and proposed major infrastructure development, most notably nuclear power-plants in the UK, in order to establish the age and cause of the largest flood over the last 8000-10,000 years.

In the USA palaeoflood analyses have gained credibility in the engineering community because of advances in statistical techniques that can incorporate non-standard observations of flood magnitude and timing (O'Connor *et al.*, 2014). FFA now uses Bayesian approaches that explicitly account for palaeohydrologic bounds as well as measurement uncertainties (O'Connell *et al.*, 2002). This is the basis of FLDFRQ3, a Bureau of Reclamation FFA program commonly used in dam safety assessments (Swain *et al.*, 2004). This program can efficiently incorporate palaeoflood, documentary and instrumental flow data, and provides more confident estimates of rare floods with annual exceedance probabilities of  $0.01-10^{-6}$ , depending on the length of the geologic record and the abundance and character of flood evidence.

## A.4 Environment Agency's approach to flood frequency analysis in the context of climate change

The Environment Agency's 2011 'Adaption to climate change: Advice for flood and coastal erosion risk management authorities' document outlines a purely model-based approach to assessing the impact of future climate change on the magnitude of river flood events up to 2080. Projections are made in relation to changes of river peak flood flows, by river basin district and compared to the 1961-1990 baseline. The baseline period is assumed to capture representative data on the magnitude of extreme flood events. This assumption on the basis of documentary and geological flood records in the UK that span periods of hundreds (Foulds and Macklin, 2015; Macdonald, 2014) or thousands (Jones *et al.*, 2012) of years is incorrect. One of the principal assumptions in the Environment Agency's approach is that peak flows in the pre-instrumental period have not been larger than the predicted  $H^{++}\%$  changes (p. 6, Table 1). Both documentary and geologically based records of extreme floods in the UK show this premise does not hold. Furthermore, arbitrary dates (p.12, figure 1) – 2039/40, 2069/70 and 2099/2100 – have been used for when modelled changes in peak flows are predicted to occur. The timing of these, however, is likely to bear no resemblance to what will happen as the result of shifts in atmospheric circulation associated with the North Atlantic Oscillation (NAO; Walker, 1924) and Atlantic Multi-decadal Oscillation (AMO; Schlesinger and Ramankutty, 1994; Kerr, 2000) which control flood frequency and magnitude in the UK (Foulds and Macklin, 2015).

From the perspective of exceptionally stringent requirements for dam or nuclear safety, particularly in the context of the uncertainty introduced by anthropogenic climate change, model extrapolations of flood risk associated with the 1-0.01% annual

probability event based on the 1961-1990 river flow series are flawed and misleading. Short (generally 30-50 years) and recent (post-1950s) gauge records are unrepresentative of future and past extreme flood events because of significant non-stationarity in the flood series over multi-decadal and longer timescales resulting from climate and catchment land-use change. The ONR will require for each nuclear power-plant site, including related communication and infrastructure links, a bespoke flood risk assessment based on long-term flood series using real data from instrumental, documentary and geological records. This will produce a 'no regrets' outcome with a credible extrapolation for a 0.01% annual probability event.

Current Environment Agency engineering and model-based approaches for extending the record of extreme river floods (1-0.01% annual probability) need to be assessed as matter of some urgency. The two fundamental assumptions of FFA – large floods are random and independent events, and stationarity of the flow series – are not met. Furthermore, arbitrary dates currently used for modelling changes in peak flows related to future climate change have no physical or process basis. It is strongly recommended that at all existing and proposed nuclear power plant sites bespoke, site-based palaeoflood investigations are undertaken to ascertain the record of extreme floods over the last 8000-10,000 years. These investigations must include an assessment of flood risk from the 0.01% annual probability event on infrastructure (road, rail and electricity network) that would be used to help evacuate personnel from the site or assist plant operators should an extreme and damaging flood occur. Palaeoflood techniques and data should also be employed in improving flood risk estimates, through flood series extension, to establish the 'probable maximum flood' in all major infrastructure development and current assets where life and property are at risk.

## **A.5 Some anticipated frequently asked questions from practitioners**

### **A.5.1 What sort of flood hydrology projects are worth the effort of palaeoflood analysis?**

All projects where there is a serious risk to life (the UK government places a value on a single life of £1.45 m at a base line date of 2000; Pavlovskaya, 2014), property and critical infrastructure, particularly where there are no gauged flow records or they are too short to accurately estimate infrequent, large events of 1-0.01% annual probability.

### **A.5.2 What sort of rivers/locations are or not suitable?**

All river catchments in the UK have historical and Holocene palaeoflood records and this report highlights where flood recording sedimentary environments can be found (see Section A.3). These are located at channel and channel margin sites (boulder berms and bars, vertical and lateral accretion units) and in overbank contexts (abandoned river channels – palaeochannels, floodbasins). Boulder berms and bars are widespread in upland river catchments; floodbasins occur in lowland as well as in lower-gradient reaches of upland and piedmont rivers; and palaeochannels are found in both upland and lowland contexts. Within heavily modified valley floors in urban or industrial areas it may be more difficult to identify alluvial landforms, but with the widespread availability of LiDAR, subtle topographic features can be mapped and likely sites for the preservation of palaeoflood deposits identified. In rural catchments potential palaeoflood sites can be easily identified from LiDAR and there are a number of automated geomorphological mapping procedures that have been developed for this purpose (e.g. Jones *et al.*, 2007). The starting point for all palaeoflood studies is an interpretative geomorphological map of the study site, or reach, showing the nature and

heights of river terraces, palaeochannels, floodbasins etc., and an independent chronology (developed from cartographic, lichenometric and/or radiometric dating) of river movement (both laterally and vertically), incision and sedimentation.

### **A.5.3 What sort of expertise and equipment is needed?**

Expertise in fluvial geomorphology, as well as hydrology, is essential for palaeoflood investigations. Other specialist skills may be required, depending on the methods of analysis needed.

Equipment requirements for palaeoflood studies, beyond the presumed availability of high resolution DEMs obtained from LiDAR or terrestrial laser scanning, include percussion corers for sampling palaeochannel fills or floodbasins. Sediment cores would need to be analysed for grain size and sampled for organic material ( $^{14}\text{C}$ ) or sand (OSL) for dating.

### **A.5.4 How do users find out about/ access existing data?**

This paper provides references to all published and publically available event-scale river palaeoflood studies undertaken in the UK at the time of writing (2015). Papers by Jones *et al.* (2010b), Lewin and Macklin (2010), Macklin and Harrison (2012), Macklin *et al.* (2012b), Foulds and Macklin (2015) provide reviews of palaeoflood techniques and data in the UK.

### **A.5.5 What sort of costs are we looking at?**

The regional and catchment-scale studies of Rumsby (1991), Merrett (2001), Jones (2007) and Foulds (2008) were all three year PhD programmes. At today's prices these equate at c. £20K per year plus coring, laboratory and dating costs of c. £5-10K. In terms of site or reach-based (1-5 km in length) studies, costs for coring, laboratory analysis and dating, and hydraulic modelling are likely to be in the region of £5K (site) - £50K (reach-scale investigation).

## References cited in Appendix A

- ARCHER, D., 1987. Improvement in flood estimates using historical flood information on the River Wear at Durham. In *Proceedings of the British Hydrological Society, First National Symposium* (Hull, January 1987), 5-1-5.9. British Hydrological Society.
- BAKER, V.R., WEBB, R.H. and HOUSE, P.K., 2002. The scientific and societal value of paleoflood hydrology. In *Ancient Floods, Modern Hazards – Principles and Applications of Paleoflood Hydrology* (ed. P.K. House, R.H. Webb, V.R. Baker and D.R. Levish), pp. 1-19. Washington, DC: American Geophysical Union.
- BENITO, G. AND O'CONNOR, J.E., 2013. Quantitative paleoflood hydrology. In *Treatise on Geomorphology, Volume 9 – Fluvial Geomorphology* (ed. E.E. Wohl and J. Shroder), pp. 459-472. San Diego: Academic Press.
- BENITO, G., THORNDYCRAFT, V.R., RICO, M., SÁNCHEZ-MOYA, Y. AND SOPEÑA, A., 2008. Palaeoflood and floodplain records from Spain: evidence for long-term climate variability and environmental changes. *Geomorphology*, 101 (1-2), 68-77.
- BENITO, G., MACKLIN, M.G., PANIN, A., ROSSATO, S., FONTANA, A., JONES, A.F., MACHADO, M.J., MATLAKHOVA, E., MOZZI, P. AND ZIELHOFER, C., 2015. Recurring flood distribution patterns related to short-term Holocene climatic variability. *Scientific Reports*, 5, Article 16398.
- BOCCALETTI, G., 2015. Stationarity. In *This Idea Must Die: Scientific Theories that are Blocking Progress* (ed. J. Brockman). New York: Harper Collins.
- BROWN, A.G., 2008. Geoarchaeology, the four dimensional (4D) fluvial matrix and climatic causality. *Geomorphology*, 101 (1-2), 278-297.
- BROWN, A.G., KEOUGH, M.K. AND RICE, J., 1994. Floodplain evolution in the East Midlands, United Kingdom: the Lateglacial and Flandrian alluvial record from the Soar and Nene valleys. *Philosophical Transactions of the Royal Society Series A*, 348 (1647), 261-293.
- BROWN, A.G., COOPER, L., SALISBURY, C.R. AND SMITH, D.N., 2001. Late Holocene channel changes of the Middle Trent: channel response to a thousand year flood record. *Geomorphology*, 39 (1-2), 69-82.
- CARLING, P.A. AND GRODEK, T., 1994. Indirect estimates of ungauged peak discharges in a bedrock channel with reference to design discharge selection. *Hydrological Processes*, 8 (6), 497-511.
- DULLER, G.A.T., 2004. Luminescence dating of quaternary sediments – recent advances. *Journal of Quaternary Science*, 19 (2), 183-192.
- ENVIRONMENTAL AGENCY, 2011. *Adapting to climate change: advice for flood and coastal erosion risk management authorities*. Bristol: Environmental Agency.
- FOULDS, S.A., 2008. *Fluvial response to late Holocene land use and climate change in the Yorkshire Dales*. Unpublished PhD thesis, Aberystwyth University.
- FOULDS, S.A. AND MACKLIN, M.G., 2016. A hydrogeomorphic assessment of twenty-first century floods in the UK. *Earth Surface Processes and Landforms*, 41 (2), 256-270.
- FOULDS, S.A., MACKLIN, M.G. AND BREWER, P.A., 2014a. The chronology and hydrometeorology of catastrophic floods on Dartmoor, south west England. *Hydrological Processes*, 28 (7), 3067-3087.

- FOULDS, S.A., GRIFFITHS, H.M., MACKLIN, M.G. AND BREWER, P.A., 2014b. Geomorphological records of extreme floods and their relationship to decadal-scale climate change. *Geomorphology*, 216, 193-207.
- FRANCIS, J.A. AND VAVRUS, S.J., 2015. Evidence for a wavier jet stream in response to rapid Arctic warming. *Environmental Research Letters*, 10 (1), 014005.
- HIGGS, G., 1987. Environmental change and hydrological response: flooding in the upper Severn catchment. In *Palaeohydrology in Practice: a River Basin Analysis* (ed. K.J. Gregory, J. Lewin and J.B. Thornes), pp. 131-160. Chichester: Wiley.
- HUDSON-EDWARDS, K.A., MACKLIN, M.G., FINLAYSON, R. AND PASSMORE, D.P., 1999. Medieval lead pollution in the River Ouse at York, England. *Journal of Archaeological Science*, 26 (7), 809-813.
- INTERAGENCY ADVISORY COMMITTEE ON WATER DATA, 1982. *Guidelines for determining flood flow frequency*. Bulletin 17B of the Hydrology Subcommittee. Reston, VA: US Geological Survey Office of Water Data Coordination.
- JONES, A.F., 2007. *Reconstructing historical and Holocene flooding in Welsh river catchments*. Unpublished PhD thesis, Aberystwyth University.
- JONES, A.F., BREWER, P.A., JOHNSTONE, E. AND MACKLIN, M.G., 2007. High-resolution interpretative geomorphological mapping of river valley environments using airborne LiDAR data. *Earth Surface Processes and Landforms*, 32 (10), 1574-1592.
- JONES, A.F., BREWER, P.A. AND MACKLIN, M.G., 2010a. Geomorphological and sedimentological evidence for variations in Holocene flooding in Welsh river catchments. *Global and Planetary Change*, 70 (1-4), 92-107.
- JONES, A.F., MACKLIN, M.G. AND LEWIN, J., 2010b. Flood series data for the later Holocene: available approaches, potential and limitations from UK alluvial sediments. *The Holocene*, 20 (7), 1123-1135.
- JONES, A.F., BREWER, P.A., MACKLIN, M.G. AND SWAIN, C.H., 2011. *Reconstructing late Holocene flood records in the Tywi catchment*. Final report for Dyfed Archaeological Trust for the Exploration Tywi! Project. Aberystwyth: Centre for Catchment and Coastal Research, University of Aberystwyth.
- JONES, A.F., MACKLIN, M.G. AND BREWER, P.A., 2012. A geochemical record of flooding on the upper River Severn, UK, during the last 3750 years. *Geomorphology*, 179, 89-105.
- JOHNSON, R.M. AND WARBURTON, J., 2002. Flooding and geomorphic impacts in a mountain torrent: Raise Beck, central Lake District, England. *Earth Surface Processes and Landforms*, 27 (9), 945-969.
- KERR, R.A., 2000. A north Atlantic climate pacemaker for the centuries. *Science*, 288 (5473), 1984-1986.
- KJELDSEN, T.R., MACDONALD, N., LANG, M., MEDIERO, L., ALBUQUERQUE, T., BOGDANOWICZ, E., BRAZDIL, R., CASTELLARIN, A., DAVID, V., FLEIG, A., GÜL, G.O., KRIAUCIUNIENE, J., KOHNOVA, S., MERZ, B., NICHOLSON, O., ROALD, L.A., SALINAS, J.L., SARAUSKIENE, D., SRAJ, M., STRUPCZEWSKI, W., SZOLGAY, J., TOUMAZIS, A., VANNEUVILLE, W., VEIJALAINEN, N. AND WILSON, D., 2014. Documentary evidence of past floods in Europe and their utility in flood frequency estimation. *Journal of Hydrology*, 517, 963-973.
- KOCHEL, R.C. AND BAKER, V.R., 1982. Paleoflood hydrology. *Science*, 215 (4531), 353-361.

- LEUSCHNER, H.H., SASS-KLAASSEN, U., JANSMA, E., BAILLIE, M.G.L. AND SPURK, M., 2002. Subfossil European bog oaks: population dynamics and long-term growth depressions as indicators of changes in the Holocene hydro-regime and climate. *The Holocene*, 12 (6), 695-706.
- LEWIN, J., 2010. Medieval environmental impacts and feedbacks: the lowland floodplains of England and Wales. *Geoarchaeology*, 25 (3), 267-311.
- LEWIN, J., 2013. Enlightenment and the GM floodplain. *Earth Surface Processes and Landforms*, 38 (1), 17-29.
- LEWIN, J. AND MACKLIN, M.G., 2010. Floodplain catastrophes in the UK Holocene: messages for managing climate change. *Hydrological Processes*, 24 (20), 2900-2911.
- LEWIN, J., MACKLIN, M.G. AND JOHNSTONE, E., 2005. Interpreting alluvial archives: sedimentological factors in the British Holocene fluvial record. *Quaternary Science Reviews*, 24 (16-17), 1873-1889.
- LLOYDS'S, 2012. *Geomorphology and changing flood risk in the UK*. London: Lloyd's Emerging Risk Team.
- LONGFIELD, S.A., 1998. *River response to recent climate and land-use change in the Yorkshire Ouse basin*. Unpublished PhD thesis, University of Leeds.
- LONGFIELD, S.A. AND MACKLIN, M.G., 1999. The influence of recent environmental change on flooding and sediment fluxes in the Yorkshire Ouse basin. *Hydrological Processes*, 13 (7), 1051-1066.
- MACDONALD, N., 2006. An underutilized resource: historical flood chronologies a valuable resource in determining periods of hydro-geomorphic change. In *Sediment Dynamics and the Hydromorphology of Fluvial Systems*, Proceedings of the IAHS/ICCE International Symposium (Dundee, July 2006), edited by J.S. Rowan, R.W. Duck and A. Werritty, IAHS Publ. 306, pp. 120-127. Wallingford: IAHS Press.
- MACDONALD, N., 2014. Millennial scale variability in high magnitude flooding across Britain. *Hydrological Earth System Sciences Discussions*, 11 (9), 10157-10178.
- MACDONALD, N. AND BLACK, A. R., 2010. Reassessment of flood frequency using historical information for the River Ouse at York, UK (1200-2000). *Hydrological Sciences Journal*, 55 (7), 1152-1162.
- MACDONALD, N., KJELDSSEN, T. R., PROSDOCIMI, I. AND SANGSTER, H., 2014. Reassessing flood frequency for the Sussex Ouse, Lewes: the inclusion of historical flood information since AD 1650. *Natural Hazards and Earth System Sciences*, 14 (10), 2817-2828.
- MACKLIN, M.G., 1999. Holocene river environments in Britain: human interaction and impact. *Journal of Quaternary Science*, 14 (6), 521-530.
- MACKLIN, M.G. AND LEWIN, J., 2003. River sediments, great floods and centennial-scale Holocene climate change. *Journal of Quaternary Science*, 18 (2), 102-105.
- MACKLIN, M.G. AND RUMSBY, B.T., 2007. Changing climate and extreme floods in the British uplands. *Transactions of the Institute of British Geographers*, 32 (2), 168-186.
- MACKLIN, M.G. AND LEWIN, J., 2008. Alluvial responses to the changing earth system. *Earth Surface Processes and Landforms*, 33 (13), 1374-1395.
- MACKLIN, M.G., RUMSBY, B.T. AND HEAP, T., 1992a. Flood alluviation and entrenchment: Holocene valley floor development and transformation in the British uplands. *Geological Society of America Bulletin*, 104 (6), 631-643.

- MACKLIN, M.G., RUMSBY, B.T. AND NEWSON, M.D., 1992b. Historic overbank floods and vertical accretion of fine-grained alluvium in the Lower Tyne Valley, north east England. In *Dynamics of Gravel-Bed Rivers*, Proceedings of the Third International Workshop on Gravel-bed Rivers (ed. P. Billi, R. Hey, P. Tacconi and C. Thorne), pp. 564-580. Chichester: Wiley.
- MACKLIN, M.G., RIDGWAY, J., PASSMORE, D.G. AND RUMSBY, B.T., 1994. The use of overbank sediment for geochemical mapping and contamination assessment: results from selected English and Welsh floodplains. *Applied Geochemistry*, 9 (6), 689-700.
- MACKLIN, M. G., JOHNSTONE, E. AND LEWIN, J., 2005. Pervasive and long-term forcing of Holocene river instability and flooding by centennial-scale climate change. *The Holocene*, 15 (7), 937-943.
- MACKLIN, M.G., JONES, A.F. AND LEWIN, J., 2010. River response to rapid Holocene environmental change: evidence and explanation in British catchments. *Quaternary Science Reviews*, 29 (13-14), 1555-1576.
- MACKLIN, M.G., FULLER, I.C., JONES, A.F. AND BEBBINGTON, M., 2012a. New Zealand and UK Holocene flooding demonstrates interhemispheric climate asynchrony. *Geology*, 40 (9), 775-778.
- MACKLIN, M.G., LEWIN, J. AND WOODWARD, J.C., 2012b. The fluvial record of climate change. *Philosophical Transactions of the Royal Society A*, 370 (1966), 2143-2172.
- MACKLIN, M.G., LEWIN, J. AND JONES, A.F., 2013. River entrenchment and terrace formation in the UK Holocene. *Quaternary Science Reviews*, 76, 194-206.
- MACKLIN, M.G., LEWIN, J. AND JONES, A.F., 2014. Anthropogenic alluvium: an evidence-based meta-analysis for the UK Holocene. *Anthropocene*, 6, 26-38.
- MAIZELS, J. AND AITKEN, J., 1991. Palaeohydrological change during deglaciation in upland Britain: a case study from northeast Scotland. In *Temperate Palaeohydrology* (ed. L. Starkel, K.J. Gregory and J.B. Thornes), pp. 105-145. Chichester: Wiley.
- MERRETT, S.P., 2001. *Historic flooding and valley floor development in the Yorkshire Dales*. Unpublished PhD thesis, University of Leeds.
- MERRETT, S.P. AND MACKLIN, M.G., 1999. Historic river response to extreme flooding in the Yorkshire Dales, northern England. In *Fluvial Processes and Environmental Change* (ed. A.G. Brown and T.M. Quine), pp. 345-360. Chichester: Wiley.
- MCEWEN, L.J., 1987. Sources for establishing a historic flood chronology (pre-1940) within Scottish river catchments. *Scottish Geographical Magazine*, 103 (3), 132-140.
- MCEWEN, L.J., 1990. The establishment of an historical historic flood chronology for the River Tweed catchment, Berwickshire, Scotland. *Scottish Geographical Magazine*, 106 (1), 37-48.
- O'CONNELL, D.R.H., OSTENAA, D.A., LEVISH, D.R. AND KLINGER, R.E., 2002. Bayesian flood frequency analysis with paleohydrologic bound data. *Water Resources Research*, 38 (5), 16-11–16-13.
- O'CONNOR, J.E., ATWATER, B.F., COHN, T.A., CRONIN, T.M., KEITH, M.K., SMITH, C.G. AND MASON, R.R., 2014. *Assessing inundation hazards to nuclear powerplant sites using geologically extended histories of riverine floods, tsunamis, and storm surges*. Scientific Investigations Report 2014–5207. Reston, VA: US Geological Survey.

- PAVLOVSKA, N., 2014. *Strategic Water Management: International Experience and Practice. Volume III: Flood risk management*. Paris: UNESCO.
- RUMSBY, B.T., 1991. *Flood frequency and magnitude estimates based on valley floor morphology and floodplain sedimentary sequences: the Tyne Basin, N.E. England*. Unpublished PhD thesis, University of Newcastle-upon-Tyne.
- RUMSBY, B., 2000. Vertical accretion rates in fluvial systems: a comparison of volumetric and depth-based estimates. *Earth Surface Processes and Landforms*, 25 (6), 617-631.
- RUMSBY, B.T. AND MACKLIN, M.G., 1994. Channel and floodplain response to recent abrupt climate change, the Tyne basin, northern England. *Earth Surface Processes and Landforms*, 19 (6), 499-515.
- SCHILLEREFF, D.N., 2015. *Lake sediment records of flood frequency and magnitude*. Unpublished PhD thesis, University of Liverpool.
- SCHILLEREFF, D.N., CHIVERRELL, R.C., MACDONALD, N. AND HOOKE, J.M., 2014. Flood stratigraphies in lake sediments: a review. *Earth Science Reviews*, 135, 17-37.
- SCHILLEREFF, D.N., CHIVERRELL, R.C., MACDONALD, N. AND HOOKE, J.M., 2015. Hydrological thresholds and basin control over paleoflood records in lakes. *Geology*, DOI:10.1130/G37261.1.
- SCHLESINGER, M.E. AND RAMANKUTTY, N., 1994. An oscillation in the global climate system of period 65–70 years. *Nature*, 367 (6465), 723-726.
- SERINALDI, F. AND KILSBY, C.G., 2015. Stationarity is undead: uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77, 17-36.
- SMITH, L.C., 2015. Stationarity. In *This Idea Must Die: Scientific Theories that are Blocking Progress* (ed. J. Brockman). New York: Harper Collins.
- SWAIN, R.E., ENGLAND, J.F., JR., BULLARD, K.L. AND RAFF, D.A., 2004. *Hydrologic hazard curve estimating procedures*. Dam Safety Research Program, Research Report DSO-04-08. Washington DC: US Bureau of Reclamation.
- TIPPING, R., 1998. The chronology of late Quaternary fluvial activity in part of the Milfield Basin, north east England. *Earth Surface Processes and Landforms*, 23 (9), 845-856.
- TOONEN, W.H.J., DE MOLENAAR, M.M., BUNNIK, F.P.M. AND MIDDELKOOP, H., 2013. Middle-Holocene palaeoflood extremes of the Lower Rhine. *Hydrology Research*, 44 (2), 248-263.
- TOONEN, W.H.J., WINKELS, T.G., COHEN, K.M., PRINS, M.A. AND MIDDELKOOP, H., 2015. Lower Rhine historical flood magnitudes of the last 450 years reproduced from grainsize measurements of flood deposits using end member modelling. *Catena*, 130, 69-81.
- TOONEN, W.H.J., MIDDELKOOP, H., KONIJNENDIJK, T.Y.M., MACKLIN, M.G. AND COHEN, K.M., 2016. The influence of long-term hydroclimatic variability on flood frequency in the Lower Rhine. *Earth Surface Processes and Landforms*, 41 (9), 1266-1275.
- TROUET, V., ESPER, J., GRAHAM, N.E., BAKER, A., SCOURSE, J.D. AND FRANK, D.C., 2009. Persistent positive North Atlantic Oscillation Mode dominated the Medieval climate anomaly. *Science*, 324 (5923), 78-80.



WALKER, G.T., 1924. Correlation in seasonal variations in weather, IX. A further study of world weather. *Memoirs of the India Meteorological Department*, 24 (9), 275-332.

WALKER, M.J.C., 1993. Holocene (Flandrian) vegetation change and human activity in the Carneddau area of upland Mid-Wales. In *Climate Change and Human Impact on the Landscape: Studies in Palaeoecology and Environmental Archaeology* (ed. F.M. Chambers), Chapter 15, pp. 169-183. London: Chapman and Hall.

WERRITTY, A., PAINE, J.L., MACDONALD, N., ROWAN, J.S. AND MCEWEN L.J., 2006. Use of multi-proxy records to improve estimates of flood risk: lower River Tay, Scotland. *Catena*, 66, 107-119.



# Appendix B: Statistical methods for the inclusion of historical data in flood frequency analysis

## B.1 Introduction

The advantages of incorporating information on historical flood events into classical flood frequency analysis (FFA) have long been recognised (Stedinger and Cohn 1986, Bayliss and Reed 2001, Payraastre et al. 2011) and a number of authors have proposed appropriate methods. The inclusion of historical records has been shown to substantially reduce the uncertainty around estimated design events and can provide insight into the rarest events which might have pre-dated relatively short systematic records of river flow. The procedures used are usually extensions of 2 estimation methods often used in hydrology: L-moments and maximum likelihood estimation.

L-moments were introduced by Hosking (1990) and are widely used by hydrologists in particular in regional flood frequency analysis (RFFA). The statistical methods described in the Flood Estimation Handbook (FEH) and its subsequent updates are based on L-moment estimation. L-moments can be computed as linear combinations of Probability Weighted Moments, and Wang (1990a) and Wang (1990b) introduced partial probability weighted moments (PPWM) to accommodate censored samples like historical records. Historical records are censored samples in the sense that a large part of the information available is actually the fact that a number of events did not exceed a certain perception threshold; for these points no information on the flow magnitude is available, but it is known that they were below a given value.

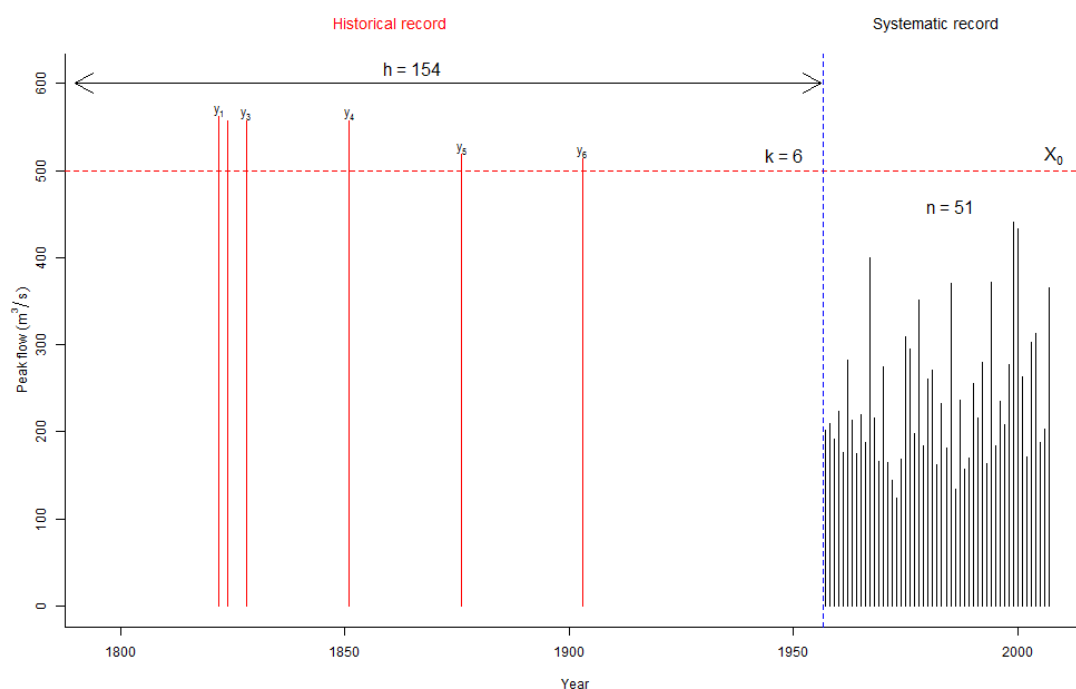
Maximum Likelihood (ML) estimation (Azzalini 1996, Coles 2001) is widely used due to its flexibility and the optimal asymptotic properties which maximum likelihood estimates possess, namely unbiasedness and efficiency. Stedinger and Cohn (1986) show how to modify the likelihood function to include historical data. In their review on methods to include historical data in flood frequency analysis, Bayliss and Reed (2001) note that approaches based on Maximum Likelihood seem to be used more frequently by researchers, despite the potential numerical failures of the maximum likelihood maximisation. Indeed, in a more recent review, Kjeldsen et al. (2014) found that, beside Spain, those countries that have standardised procedures in place for the use of historical data in FFA recommend the use of Maximum Likelihood approaches that can also be combined with a Bayesian approach. A brief introduction to (Partial) Probability Weighted Moments and Maximum Likelihood methods is given below. The reader is referred to the references in the text for more information.

The standard methods for flood frequency estimation rely on samples of systematic records of measured high flow at a given gauging station, represented as  $\mathbf{x} = (x_1, \dots, x_n)$ . Typically, it is assumed that the data available follow a specific probability distribution indexed by some parameters  $\theta$  ( $X \sim F(x, \theta)$ ), and statistical methods are employed to estimate the parameters of the distribution based on the available data. When historical data are available, typically this corresponds to the information that some floods of a certain magnitude occurred at a point in time and that these floods correspond to the biggest events in a certain range of time (for example, since the beginning of the printing of a local newspaper). In particular, the methods presented in this appendix deal mostly with the case in which the magnitude of  $k$  events across  $h$

years is known, although some investigation of the case in which only the fact that some  $k$  event exceeded the perception threshold is also pursued.

All  $k$  events have a magnitude above the value,  $X_0$ , which is often named the perception threshold, since it corresponds to the threshold above which the flood would have been large enough to be noted in historical sources or to leave recognisable signs across the catchment. One important assumption made in this setting is that the  $k$  historical floods for which some information is available correspond to all the events above the perception threshold which have happened in the period of time covered by the  $h$  years.

A good understanding of what values correspond to  $h$ ,  $k$  and  $X_0$  is a necessary prerequisite before applying any estimation procedure that combines systematic and historical data, as these values would have a large impact on the final estimates (see, for example, the discussion in Strupczewski et al. 2014 or Macdonald et al. 2014). Figure B.1 shows an exemplification of the quantities  $h$ ,  $k$ ,  $n$  and  $X_0$  used throughout the report.



**Figure B.1 Historical data example (River Wear at Durham), showing a total of  $k = 6$  historical events (red bars) above the perception threshold  $X_0$  (dashed red line), recorded across the  $h = 154$  year-long historical period**

Notes: The  $n = 51$  years' long systematic record of gauged peak flows is also shown (black bars).

One additional important aspect of any statistical procedure used to estimate flood frequencies is that it is assumed that all data points, both from the systematic and the historical records, come from the same distribution, that is, that the process under study is actually stationary. If there is reason to believe that large changes have occurred in the flood generation process (for example, changes in the basin properties, disruptions of the floodplain or the river channel that could alter the hydraulic properties of the river), a thorough assessment of whether events from the past can be representative of the present situation should be performed. This is also valid for changes that might be the results of climate change, for example, diminishing snowfall and snowmelt.

Payraastre et al. (2011) note that changes in the basin properties are not likely to affect the magnitude of extremely large and rare events, and it is likely that including historical events would still give a more complete information on the very rare cases. For events of relatively small size (for example, a magnitude is line with those recorded in a 25-year record), it is important to ensure that the historical record can be directly integrated with the systematic records.

Given the fluctuations between flood rich and flood poor periods, information about past events could help to give better estimates of the frequency of large events which might have not been registered in the systematic record. Ideally, historical records could be used to gain better understanding of the natural fluctuations of river flows, as hinted in Macdonald (2014). In the likelihood approach, a non-stationary model could be employed to relate flood risk to one or more external variables, but a very long and rich historical record would be required to gain a good understanding of large-scale variabilities.

## B.2 Methods

### B.2.1 (Partial) Probability Weighted Moments

Wang (1990a) and Wang (1990b) introduced partial probability weighted moments (PPWM), an extension of the Probability Weighted Moments (PWM). These can easily accommodate the use of both historical and systematic data in the estimation of the parameters of a distribution. As discussed by Hosking and Wallis (1997) and the references therein, PWM estimation methods are equivalent to L-moment estimation methods, since L-moments can be written as linear combinations of the PWM.

The probability weighted moments of a random variable  $X$  with distribution function  $F(x) = P(X \leq x)$  are defined as:

$$M_{p,r,s} = \int_0^1 [x(F)]^p F^r (1-F)^s dF$$

(equation B.1)

where  $p, r$  and  $s$  are real numbers. When  $p = 1$  and  $s = 0$ , the moments become:

$$\beta_r = M_{1,r,0} = \int_0^1 x(F) F^r dF$$

(equation B.2)

and, given a sample of data points  $(x_{(1)}, \dots, x_{(n)})$  ordered from the smallest  $(x_{(1)})$  to the largest  $(x_{(n)})$ , unbiased estimators of  $\beta_r$  can be found as:

$$b_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2) \dots (i-r)}{(n-1)(n-2) \dots (n-r)} x_{(i)}.$$

(equation B.3)

Partial Probability Weighted Moments, which can be used to estimate the distribution underlying a sample with, for example, lower bound censoring. PPWM for lower bound censoring at a lower threshold  $x_0$  are defined by Wang (1990a) as:

$$M'_{p,r,s} = \int_{F_0}^1 [x(F)]^p F^r (1-F)^s dF$$

(equation B.4)

where  $F_0 = F(X_0)$  corresponds to the value of the distribution at the point  $X_0$ . Unbiased estimates for  $\beta'_r = M'_{1,r,0}$  are given by:

$$b'_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2) \dots (i-r)}{(n-1)(n-2) \dots (n-r)} x_{(i)}^*$$

(equation B.5)

where:

$$x_{(i)}^* = \begin{cases} 0 & : x_{(i)} \leq x_0 \\ x_{(i)} & : x_{(i)} > x_0 \end{cases}$$

(equation B.6)

The definition under upper bound censoring PPWM  $\beta''_r$  is derived by Wang (1990a) in a similar fashion. Its estimate  $b''_r$  is calculated as:

$$b''_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2) \dots (i-r)}{(n-1)(n-2) \dots (n-r)} x_{(i)}^*$$

(equation B.7)

where:

$$x_{(i)}^* = \begin{cases} x_{(i)} & \text{if } x_{(i)} \leq X_0 \\ 0 & \text{if } x_{(i)} > X_0 \end{cases}$$

(equation B.8)

Finally, the PWM can be rewritten as the sum of the upper bound censored and lower bound censored PPWMs:

$$\beta_r = \int_0^1 x(F) F^r dF = \int_0^{F_0} x(F) F^r dF + \int_{F_0}^1 x(F) F^r dF = \beta''_r + \beta'_r$$

(equation B.9)

and estimated as the sum of the 2 PPWM estimates:  $b_r = b'_r + b''_r$ .

In the case in which historical data are used, the  $F_0 = F(X_0)$  values correspond to the percentile corresponding to the perception threshold  $X_0$ . Values below the perception threshold are known only for the systematic record of length  $n$  and these are used to estimate the  $\beta'' = \int_0^{F_0} x(F) F^r dF$ :

$$b''_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2) \dots (i-r)}{(n-1)(n-2) \dots (n-r)} x_{(i)}^*$$

(equation B.10)

where:

$$x_{(i)}^* = \begin{cases} x_{(i)} & \text{if } x_{(i)} \leq X_0 \\ 0 & \text{if } x_{(i)} > X_0 \end{cases}$$

(equation B.11)

Further, it is known that a total of  $k$  events have exceeded  $X_0$  during the historical record and an additional  $k_s \leq n$  values in the systematic record might have exceeded the perception threshold, and conversely  $(h + n - k - k_s)$  events are known to not have exceeded the threshold. Hence, there is a total of  $k + k_s$  events which exceeded the threshold over the period of length  $(h + n)$ . Taking the ordered sample of events above the threshold  $x_{(h+n-k-k_s+1)}^h \leq x_{(h+n-k-k_s+2)}^h \leq \dots \leq x_{(h+n)}^h$ , estimate of  $\beta' = \int_{F_0}^0 x(F) F^r dF$  are obtained as:

$$b'_r = \frac{1}{h+n} \sum_{i=h+n-k-k_s+1}^{h+n} \frac{(i-1)(i-2) \dots (i-r)}{(h+n-1)(h+n-2) \dots (h+n-r)} x_{(i)}^h$$

(equation B.12)

Finally, estimates for the probability weighted moments are taken as the sum of the two partial probability weighted moments  $b_r = b'_r + b''_r$ . Once estimates for the Probability Weighted Moments are obtained, the sample L-moments are computed as linear combination of the  $b_r$  estimates and from these the parameters of any distribution are estimated.

## B.2.2 Maximum Likelihood

Given a sample of observed data  $\mathbf{x}$ , assumed to be independent and identically distributed with distribution  $f(x, \theta)$ , maximum likelihood estimates for the distribution parameter(s)  $\theta$  correspond to the values of  $\theta$  which maximises the likelihood function defined as:

$$L(\theta, \mathbf{x}) = \prod_{i=1}^n f(x_i, \theta)$$

(equation B.13)

where  $f(x_i)$  indicates the probability density function of the assumed distribution.

Under some conditions on the likelihood function, maximum likelihood estimates enjoy some asymptotic properties which are desirable in an estimate, such as consistency, unbiasedness and efficiency (for example, minimal variance). Combined with the asymptotic normality of the maximum likelihood estimates, these properties are one of the main reasons behind the wide usage of the maximum likelihood estimation. Furthermore, maximum likelihood theory gives a direct method to quantify the uncertainty around parameter estimates. Uncertainties around transformations of the parameters, such as the quantile function used for the estimation of flood frequency curves, can be obtained using the delta method.

If both historical and systematic data are available in a flood frequency analysis estimation, the likelihood is constructed by combining the information on the  $n$  systematic data points with the information that in  $h$  years a total of  $k$  events have exceeded the threshold (Stedinger and Cohn 1986). The number of threshold exceedances,  $k$ , is modelled as a binomial random variable  $K \sim \text{Bin}(h, p_0)$ , where  $p_0$  corresponds to the probability that the maximum flow in a year exceeds the perception threshold:

$$p_0 = [1 - F_x(X_0)] \quad (\text{equation B.14})$$

Further, by virtue of the law of total probability and considering that the value of an historical peak flow is known only if it is higher than the perception threshold  $X_0$ , the distribution of a historical peak flow value  $y_j$  can be shown to be:

$$\begin{aligned} f(y_j, \theta) &= f(y_j, \theta | y_j > X_0)(1 - F_x(X_0)) + f(y_j, \theta | y_j \leq X_0)F_x(X_0) = \\ &= f(y_j, \theta | y_j > X_0)(1 - F_x(X_0)) \quad (\text{equation B.15}) \end{aligned}$$

After some reworking, as shown in Macdonald et al. (2014), the likelihood function can finally be written as:

$$L(\theta, \mathbf{x}, \mathbf{y}) = \prod_{i=1}^n f(x_i, \theta) * \prod_{j=1}^k f(y_j, \theta) * \binom{h}{k} F(X_0, \theta)^{h-k}$$

(equation B.16)

where  $\mathbf{y} = (y_1, \dots, y_k)$  indicate the  $k$  values of the historical events above the threshold.

In the case in which only the information that the threshold  $X_0$  is exceeded  $k$  times is available and no information can be retrieved on the magnitude of the event, the likelihood becomes:

$$L(\theta, \mathbf{x}, \mathbf{y}) = \prod_{i=1}^n f(x_i, \theta) * \binom{h}{k} F(X_0, \theta)^{h-k} (1 - F(X_0, \theta))^k$$

(equation B.17)

The likelihood approach could also potentially accommodate the case in which different perception thresholds are exceeded in different periods of the historical record and can also accommodate for the case in which an interval of plausible event magnitude is given, rather than a point estimate. These extensions are not investigated in detail here, but they are a testimony to the great flexibility of likelihood-based methods.

In spite of the widespread use of L-moments in flood frequency analysis, in particular in Regional Flood Frequency Analysis (RFFA) (Hosking and Wallis 1997), it appears that most of the work on methods to include historical data focus on maximum likelihood estimation procedures (Stedinger and Cohn 1986), in particular the Bayesian implementation of such procedures (see, for example, Reis and Stedinger 2005, Gaume et al. 2010). It seems that the initial work of Wang (1990a) and Wang (1990b) on partial probability weighted moment (PPWM) has not been applied much, although the application does not seem to be particularly complicated. To extend the statistical methods implemented in Institute of Hydrology (1999) and subsequent updates, it would seem natural to develop methods that are in line with L-moments to augment systematic records with historical data.

The potential drawback of using PPWM and similar methods is that they do not allow for flexibility in terms of different type of data entering the model of likelihood-based methods and that, unlike likelihood-based methods, they do not allow for a direct and straightforward estimation of the uncertainty of the estimated frequency curve. However, the maximisation of the likelihood functions in the maximum likelihood methods is generally done via numerical methods and these can give unstable results and require the user to intervene to find and verify the final estimates. When a Bayesian approach to maximum likelihood is chosen, these checks need to be even more accurate and time-consuming.

A comparison of the performance of the (P)PPWM and maximum likelihood methods based on a large simulation study is given below. Only the traditional approach to



maximum likelihood, rather than Bayesian methods, is employed in this project. Bayesian methods can provide a greater flexibility in the modelling approach and a more direct estimation of the uncertainties around specific quantities (for example, design events). Nevertheless, their use is less straightforward than the classical maximum likelihood approach, and a careful check on the convergence of the Markov chain Monte Carlo (MCMC) algorithm should be made before using its outputs as basis for any statistical inference. Given the large number of simulations carried out in this study, it is not feasible to actually verify the convergence of the MCMC algorithm for all simulated samples and it is expected that the performance of the Bayesian approach with non-informative priors would be comparable with the one of the traditional maximum likelihood method.

## B.3 Simulation study settings

The simulation study is designed to be informative about the possible outcomes of including historical data in FFA for British catchments. Its aims are to:

- compare the sample properties of estimates obtained using the maximum likelihood-based methods and the Partial Probability Weighted Moments (PPWM).
- quantify the potential benefit of including historical data in the estimation process

For each simulation setting, synthetic flood data following a GLO distribution were generated. The location parameter was taken to be  $33\text{m}^3/\text{s}$  (the median location parameter for the British series of annual maxima) and the L-CV was kept constant at 0.2 (the median L-CV for British records). The shape (or skewness) parameter was taken to be a value in the set  $(-0.3, -0.1, 0.1)$ ; most AMAX series exhibit a negative shape parameter (and hence a lower bounded distribution with no upper bound), but the case of a positive shape parameter (and hence an upper bounded distribution) was included to have a more general simulation setting. The choice of location parameter should have little impact on the overall performance of the estimation procedures and the L-CV is constant across all simulations.

The length of the systematic record was taken to be a value among  $(10, 30, 46, 76)$  to be representative of very short and fairly long flow series. The length of the historical record was taken to be the length of the systematic record multiplied by a constant among  $(0, 0.5, 1, 2, 5, 10, 20, 50)$ , with 0 corresponding to the case in which only systematic data are used in the estimation procedure.

Finally  $X_0$ , the perception threshold above which historical data are included in the inference procedure, was given values corresponding to the  $(0.85, 0.9, 0.95, 0.99)$  left percentile of the underlying distribution, for example, approximately the 6.7-year, 10-year, 20-year and 100-year event. To make the simulation setting more realistic, for the case in which the historical record is 20 or 50 times longer than the historical record, no simulation using the 6.7-year and 10-year threshold was performed.

A total of 288 simulation settings were included in the study to investigate how different properties of the data included in a statistical analysis affect the final quality of the estimation procedure. While some of the overall combinations of historical and systematic record lengths and perception threshold are likely to be a fairly uncommon in reality, they were included for completeness.

For each setting, a total of  $S = 10,000$  samples were generated with the assigned systematic record length and at least one historical event exceeding the assigned

threshold; all samples were analysed using both likelihood-based methods and (P)PWM.

The quality of the estimation procedures is compared by looking at some summary measures calculated based on the estimates obtained in the  $S = 10,000$  simulations, namely the Bias, the Standard Error (SE) the Root Mean Square Error (RMSE) of the distribution parameters and the logarithm of  $Q_T$  values across a set of return periods:

$$Bias(\theta) = \frac{1}{S} \sum_{i=1}^S (\hat{\theta}_i - \theta)$$

(equation B.18)

$$SE(\theta) = \sqrt{\frac{1}{S-1} \sum_{i=1}^S (\hat{\theta}_i - \bar{\hat{\theta}})^2}$$

(equation B.19)

$$RMSE(\theta) = \sqrt{\frac{1}{S} \sum_{i=1}^S (\hat{\theta}_i - \theta)^2}$$

(equation B.20)

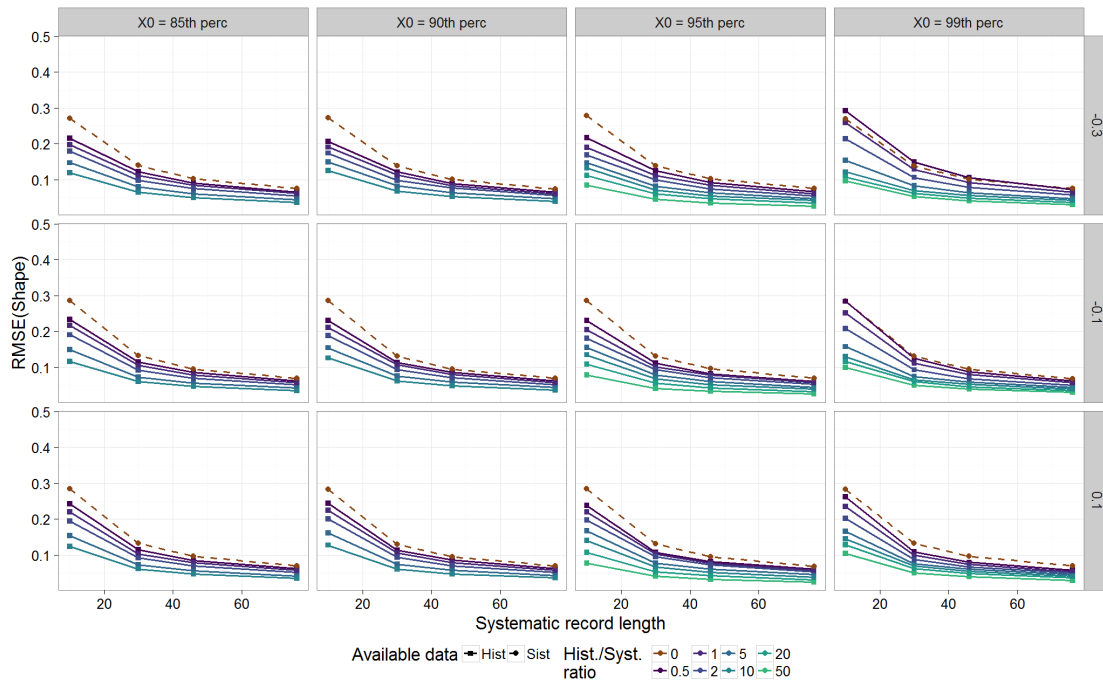
Ideally, an estimate should be unbiased (Bias = 0) and have the smallest possible standard error. The difference between the standard error and the RMSE is that the former gives an indication of how much the value of the estimate varies around the average estimated value, while the latter gives an indication of how variable are the estimates compared to the true value of the parameter. If an estimation procedure gives a biased estimate, the sample of estimated values might have a small variability around the wrong value of the parameter estimate and hence give a small standard error, but a large RMSE. For unbiased estimators, the 2 values should correspond.

## B.4 Parameter and design events estimation

This section gives an overview of the performance of the different estimation methods. The performance of the methods is assessed mostly by comparing the summary quantities (Bias, SE and RMSE) across the different simulation settings for the most important parameters and the estimated resulting quantiles.

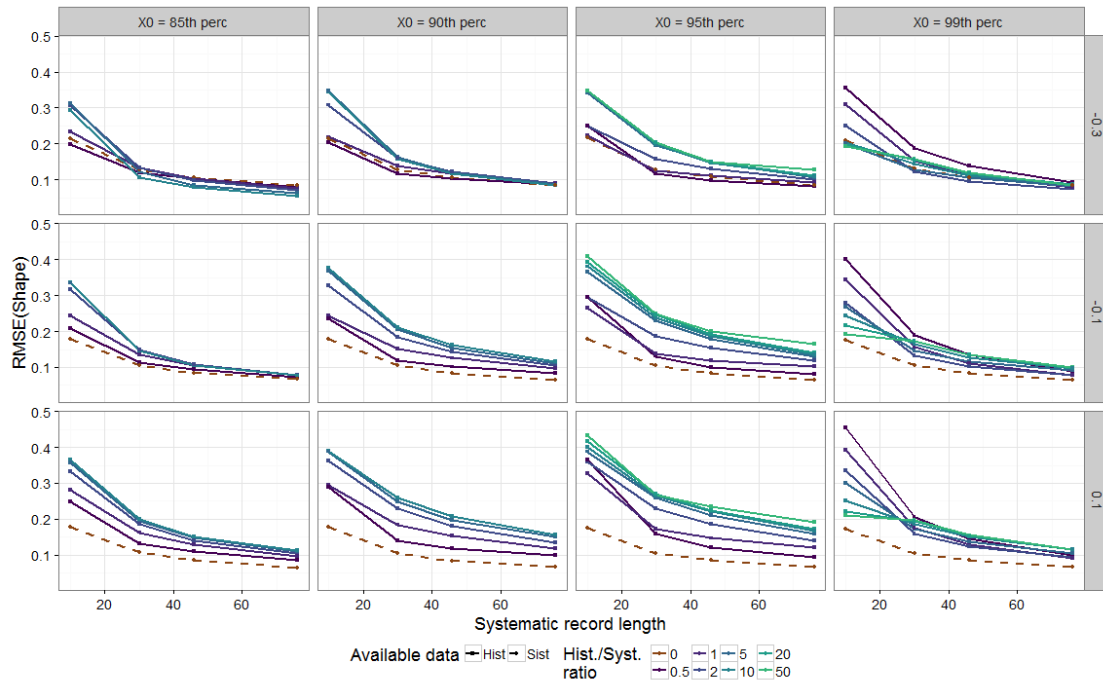
The results for the shape parameter are investigated first, as this parameter corresponds to the most uncertain and yet important for the estimation of events with long return periods. Results for the other parameters follow similar patterns and are not presented.

Figure B.2 and Figure B.3 show the RMSE values for the shape parameter for all the simulation settings obtained using the likelihood and the (P)PWM methods, respectively. In each plot, the different colours of each line indicate a specific ratio of historic to systematic record length, including the systematic case for which the ratio is 0.



**Figure B.2 RMSE for the shape parameter as a function of the systematic record length, for the different historical record length using the likelihood estimation method**

Notes: Each panel shows different  $X_0$  and shape parameter combination.



**Figure B.3 RMSE for the shape parameter as a function of the systematic record length, for the different historical record length using the (P)PWM estimation method**

Notes: Each panel shows different  $X_0$  and shape parameter combination

In both approaches, the RMSE decreases for longer systematic records, but while including historical data in the likelihood approach consistently brings a decrease in RMSE, this does not happen for the (P)PWM approach (Figure B.3). This is partially in

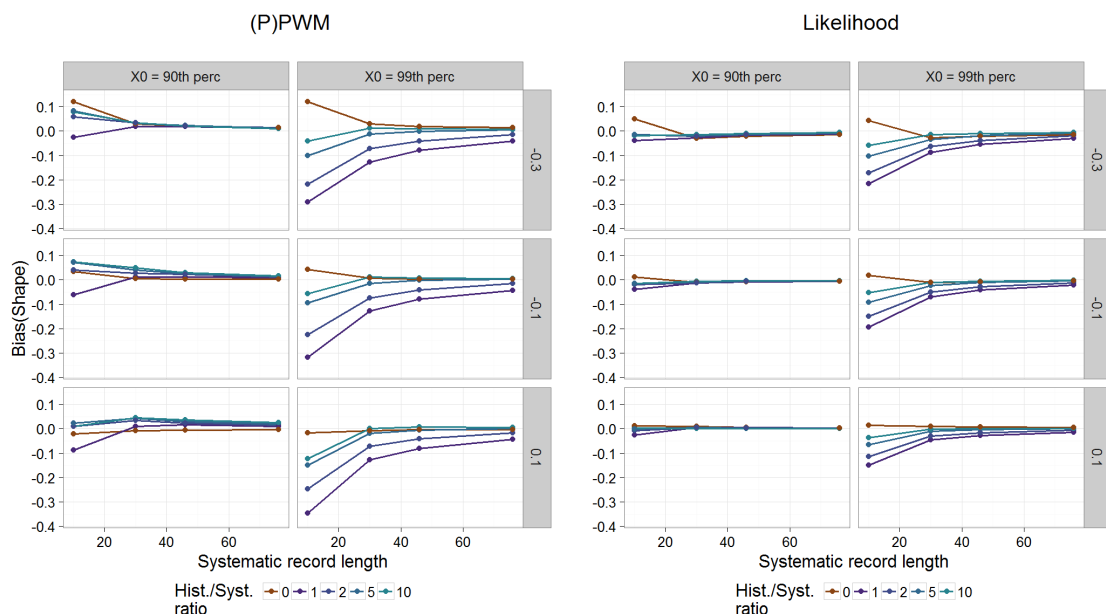
contrast with the results of the simulation study in Wang (1990b), where a decrease in RMSE is shown for a GEV with location parameter 0, scale parameter 1 and shape parameter chosen among (-0.2, 0, 0.2), corresponding to L-CV values of (1.05, 1.20, 1.45).

From other simulations not shown here, it would seem that the (P)PWM approach would consistently have lower RMSE when historical data are included in the statistical analysis when applied to series that are generated from distributions with a high L-CV, like the ones used in Wang (1990b). For cases investigated in this study, with a relatively low L-CV (0.2), including historical data in a (P)PWM estimation was found to bring lower RMSE only in few cases and only for the lowest shape parameter value (for example, -0).

When using the likelihood approach, the largest drops in RMSE are obtained when the perception threshold is relatively low (that is, more points from the historical period are available) and, when high perception threshold are used, for positive shape parameters. One further remark on comparing the results in Figure B.2 and Figure B.3 is that, when systematic data only are used, the RMSE values for the PWM method tend to be lower than the ones for the likelihood method, more so for short records (10-year and 30-year systematic record).

While this is not a surprising result, it confirms the reasons behind the wide popularity of L-moments, which stem from PWM, in hydrological science where the available record are often not much longer than 40 years. Interestingly, for some cases, even when historical data covering a relatively short period of time are integrated with a short systematic record in the likelihood approach, the PWM approach on the systematic record only delivers lower RMSE.

To further investigate the performance of the estimation procedure, Figure B.4 and Figure B.5 show the Bias and Standard Error, respectively, for the shape parameter for selected  $X_0$  perception threshold and historic to systematic record length ratios. The results for the (P)PWM and the likelihood approach are placed next to each other to aid the comparison.

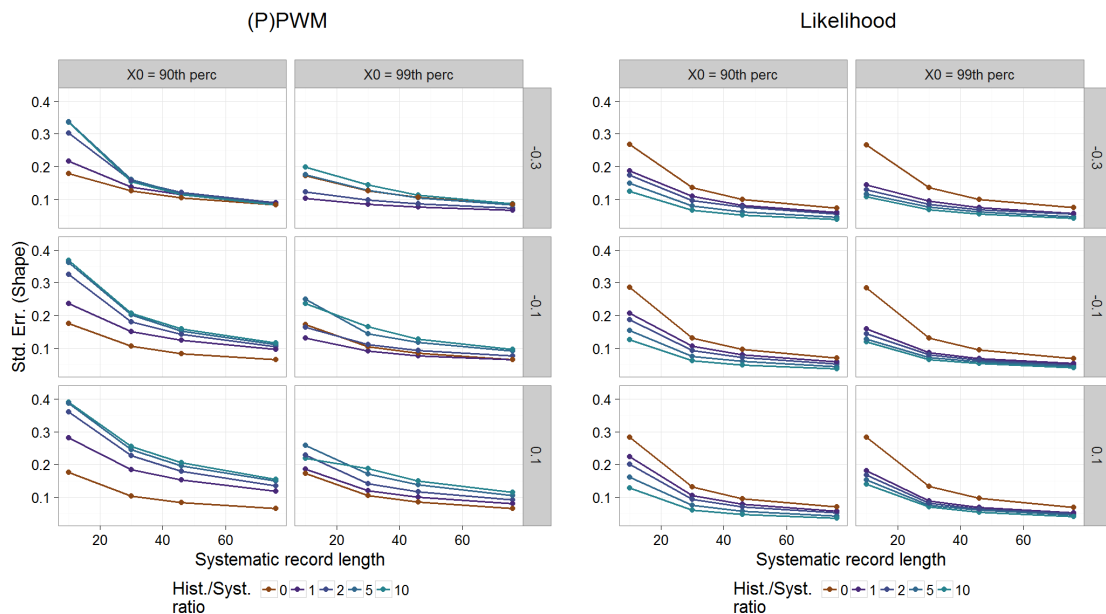


**Figure B.4 Bias for the shape parameter as a function of the systematic record length, for selected ratio of historical to systematic record lengths**

Notes: Each panel shows a different  $X_0$  and shape parameter combination.

It is noticeable in Figure B.4 that, when a very high perception threshold is used (the 100-year event), both methods give biased results for shape parameter, in particular for the case in which the systematic record is only 10 years long. For short records, however, the bias for the shape parameter increases the more negative the shape parameter gets when systematic data only are used. The reason behind this behaviour is that, for short records, it is very unlikely to record a very high peak. Therefore, all methods will tend to estimate a more positive shape parameter than the true one, thus estimating flatter flood curves. Since the more negative the shape parameter (for example, the larger the absolute value of a negative parameter) the steeper the flood frequency curve, it is not surprising that flood frequency curves based on relatively short records would not need to accommodate for very large and rare data points. However, if any data point in an historical record of less than 100 years is larger than the 100-year event, the estimated shape parameter is likely to be smaller (for example, bigger in absolute value) than the true one, as the final fitted model would need to be steeper than the one of the true model to fit the unlikely point.

In a real case scenario, one would not know to which percentile of the statistical distribution of the flood generating process the perception threshold corresponded, so it is possible that combining a short record with a very large historical event would result in biased estimates for the shape parameter. This demonstrates the importance, when investigating the available historical data for a location, of the accurate collection of information on the longest possible series of past records and on as many flood events as possible to avoid introducing large biases in the estimation –, especially if the systematic record is not very long.

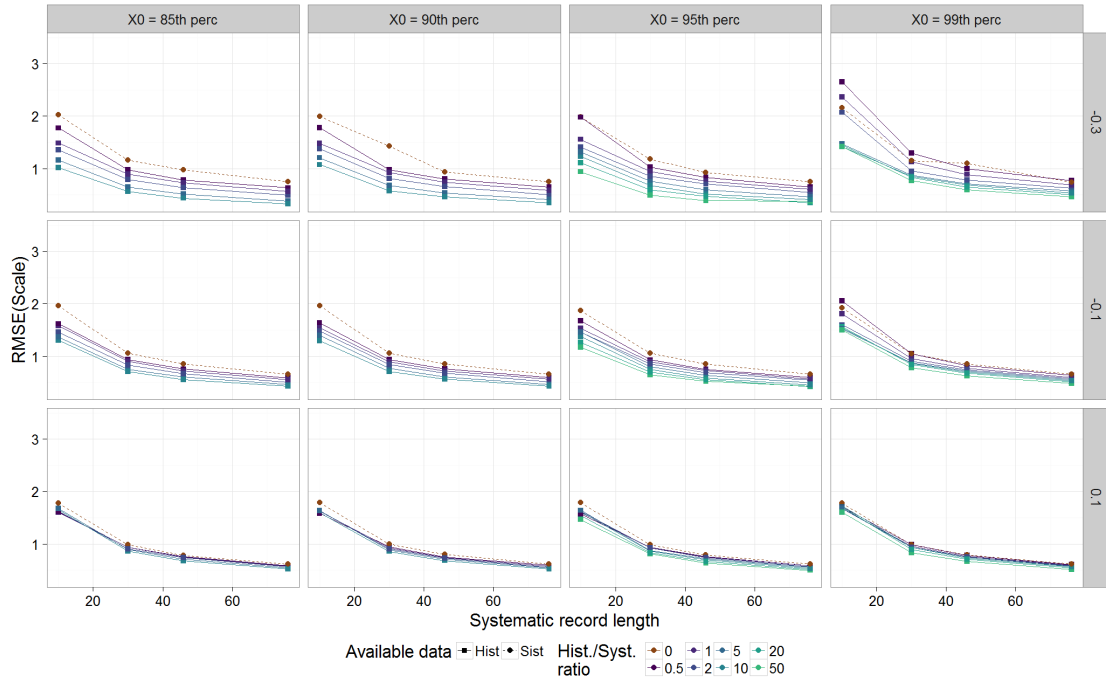


**Figure B.5 Standard error for the shape parameter as a function of the systematic record length, for selected ratio of historical to systematic record lengths**

Notes: Each panel shows a different  $X_0$  and shape parameter combination.

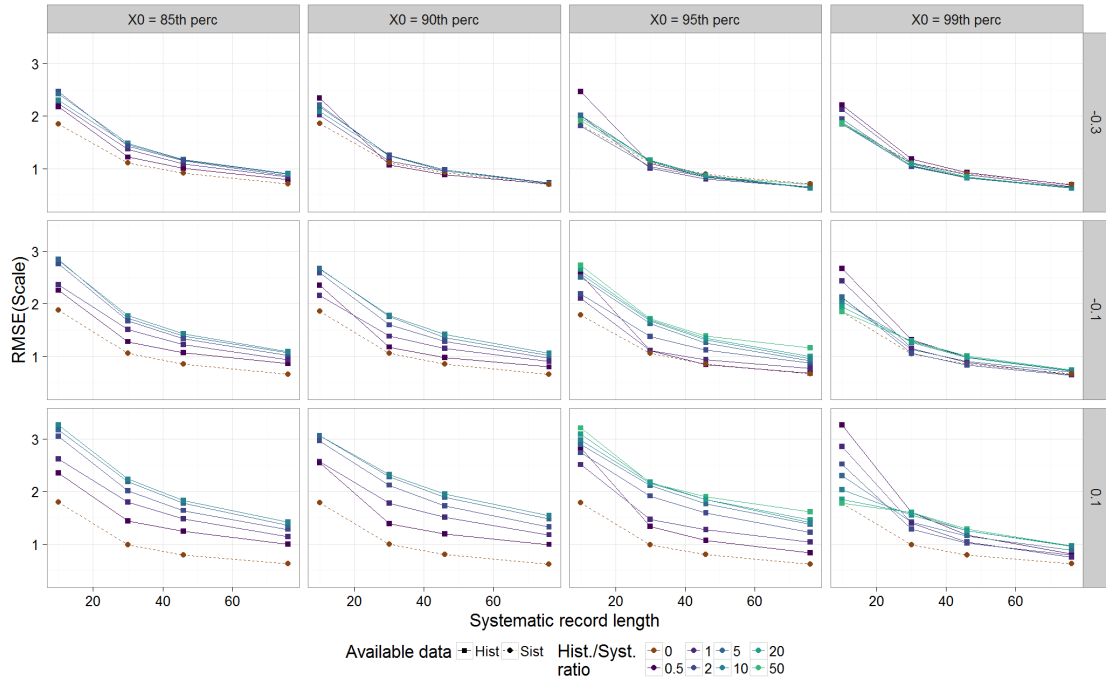
The results in Figure B.5 for the standard error values are comparable to the ones seen for the RMSE values: the likelihood method always improves when historical data are added to the estimation procedure, but when systematic data only are used in the estimation procedure, the PWM method gives much lower standard error values than the likelihood approach.

To further investigate the overall performance in terms of the parameter estimation of the different methods under different settings, Figure B.6 and Figure B.7 show the RMSE values for the scale parameter estimates for all the simulation settings obtained using the likelihood and the (P)PWM methods, respectively.



**Figure B.6 RMSE for the scale parameter as a function of the systematic record length, for the different historical record length using the likelihood estimation method**

Notes: Each panel shows different  $X_0$  and shape parameter combination.



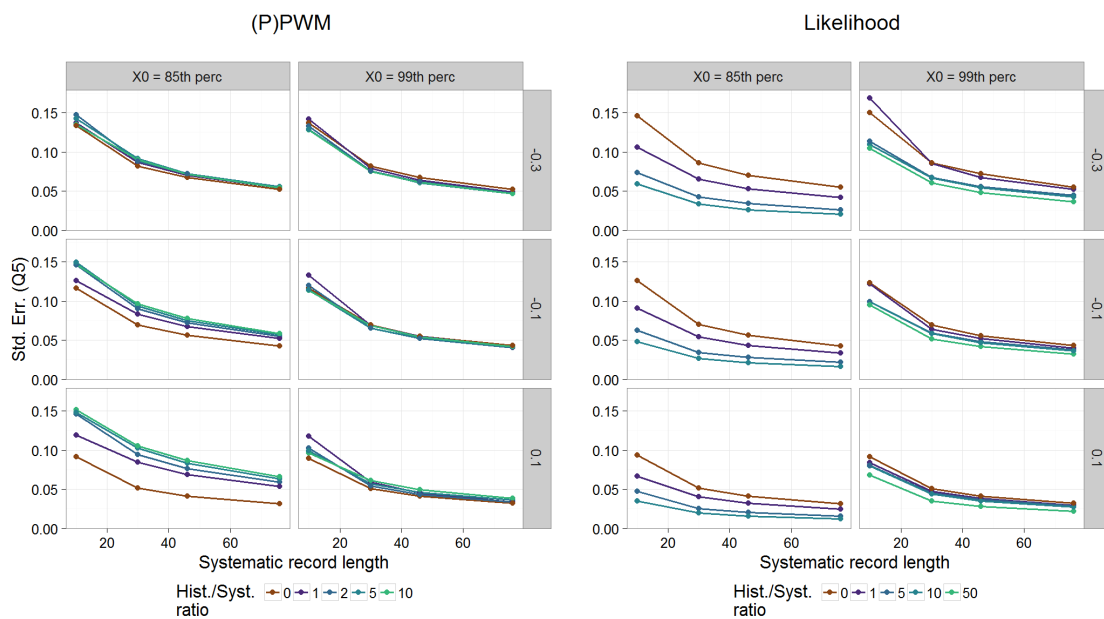
**Figure B.7 RMSE for the scale parameter as a function of the systematic record length, for the different historical record length using the (P)PWM estimation method**

Notes: Each panel shows different  $X_0$  and shape parameter combination

The overall patterns for the scale estimation is fairly similar to the one observed for the estimation of the skewness parameter. Larger sample sizes correspond to lower RMSE and the inclusion of historical data is more beneficial in the maximum likelihood approach rather than when using the (P)PWM approach.

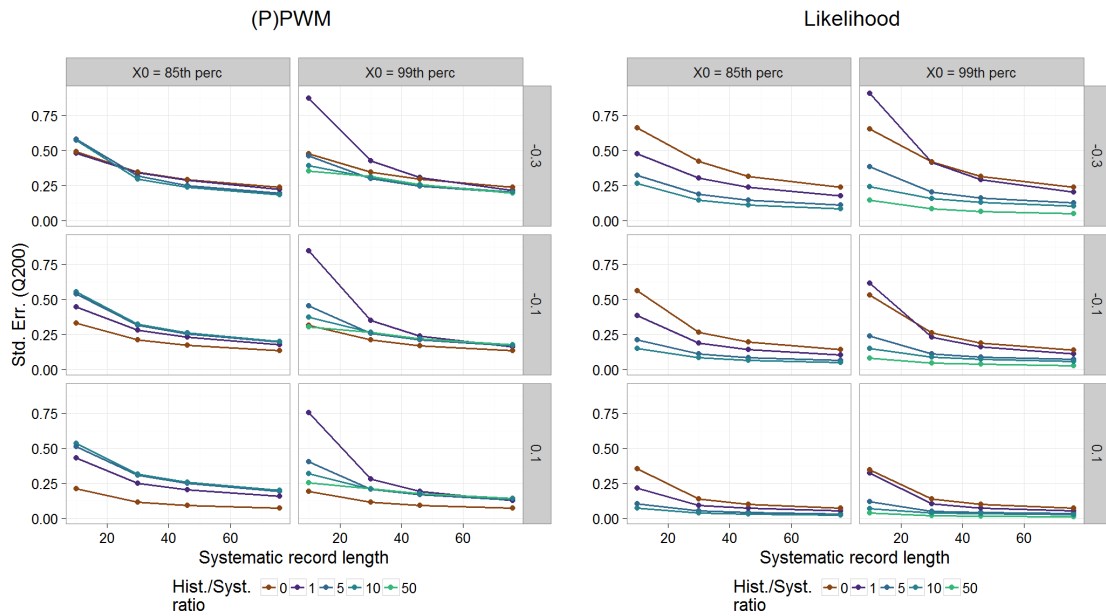
Finally, since the final aim of flood frequency analysis is the estimation of some specific quantiles of the distribution, the performance of the different methods in the various settings is investigated. The properties of the estimates for the distribution quantile are influenced by the estimation of the 3 parameters in a non-linear fashion and it is therefore a useful summary to investigate the estimation of some specific quantiles. In particular, the properties of the RMSE for a low and high quantile (namely the 5-year and the 200-year event) are investigated here – shown for selected perception threshold in Figure B.8 and Figure B.9. The chosen quantiles correspond to the events which would be exceeded with probability 0.2 and 0.005 in any given year. The RMSE values are calculated on the log-quantiles rather than on the original scale. Each figure shows results for both the (P)PWM and the likelihood approach to facilitate direct comparison.

For both quantiles, the likelihood approach gives a better framework for the inclusion of historical data, with longer historical periods and lower perception thresholds giving larger reductions in the RMSE. Not surprisingly, the RMSE values for the Q200 estimates are much larger than the ones for the Q5 estimates; higher uncertainties can be expected for higher quantiles.



**Figure B.8 RMSE for Q5 as a function of the systematic record length for selected ratio of historical to systematic record lengths**

Notes: Each panel shows different  $X_0$  and shape parameter combination



**Figure B.9 RMSE for Q200 as a function of the systematic record length for selected ratio of historical to systematic record lengths**

Notes: Each panel shows different  $X_0$  and shape parameter combination

## B.5 Uncertainty of flood frequency curves

The key qualities expected from an estimate are unbiasedness and minimal variance. That is to say it is desirable for an estimate to be, on average, equal to the true value of the quantity under study and to have little variability around this value.

Traditionally, the output of a FFA corresponded to a series of estimated quantiles corresponding to the events that are expected to be exceeded with some pre-specified probability. However, it is more and more important to also understand how variable (that is, uncertain) the estimated quantiles are. That is to say, not only it is important to provide an estimate as correct as possible for the quantiles of interest, but it is also important to be able to be confident that the true value of the quantity of interest might be not too far from the given estimate.

Factorial standard errors (fse) are often used to quantify uncertainty around the estimated quantiles. The justification behind their use of the fse stems from the assumption that the quantile values are assumed to be log-normally distributed ( $Q \sim LN(\mu, \sigma)$ ). This would mean that a 95% confidence interval for the expected value of  $\log(Q)$  could be obtained as  $(\hat{\mu} - 1.96\hat{\sigma}, \hat{\mu} + 1.96\hat{\sigma})$ , with  $\hat{\mu}$  and  $\hat{\sigma}$  appropriate estimates for  $\mu$  and  $\sigma$ . An approximate 95% confidence interval for the expected value of the original  $Q$  could then be obtained as  $(\exp(\hat{\mu})/(1.96\exp(\hat{\sigma})), \exp(\hat{\mu})(1.96\exp(\hat{\sigma})))$ . The fse corresponds to  $\exp(\hat{\sigma})$  and can be used to quantify the uncertainty around an estimate: lower fse values would indicate a smaller range of values in which one would have some confidence the true value of the parameter lies.

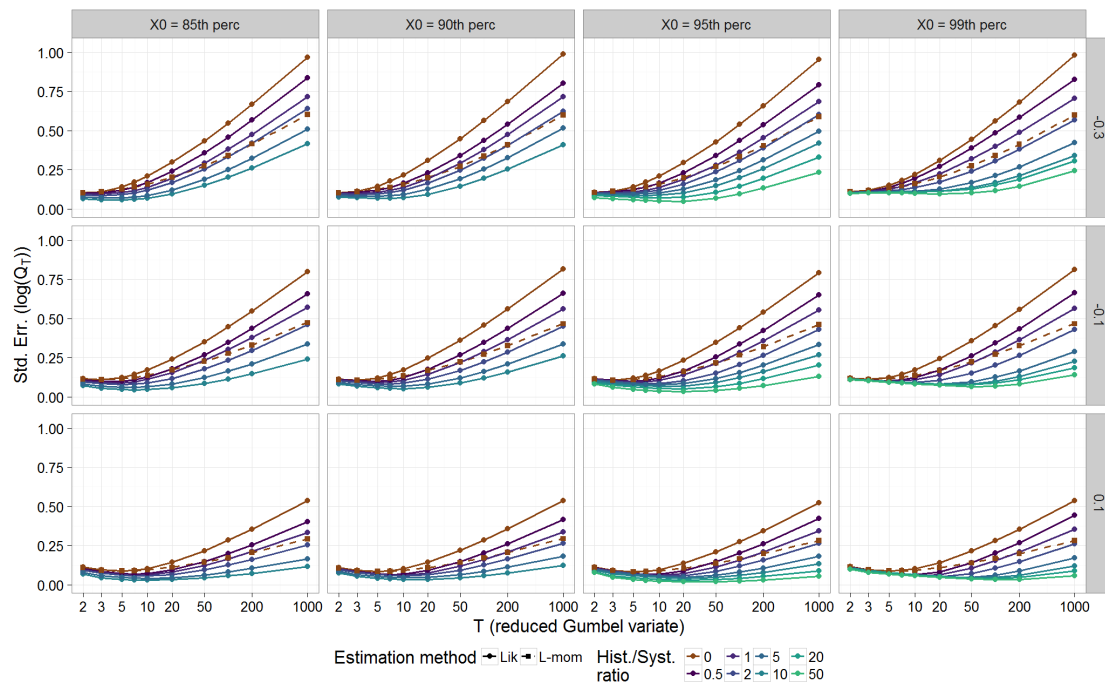
In this section, the uncertainty around the estimation of the distribution quantiles, which correspond to the design events, is investigated. To make the figures more readable, the comparison between the uncertainties under different models is made using the standard error of the estimates, which corresponds to the logarithm of the fse values.

For each simulated sample, estimates for the event that is expected to be exceeded on average every  $T$  years, that is, the  $T$ -year event, is estimated as the  $p^{th}$  percentile of



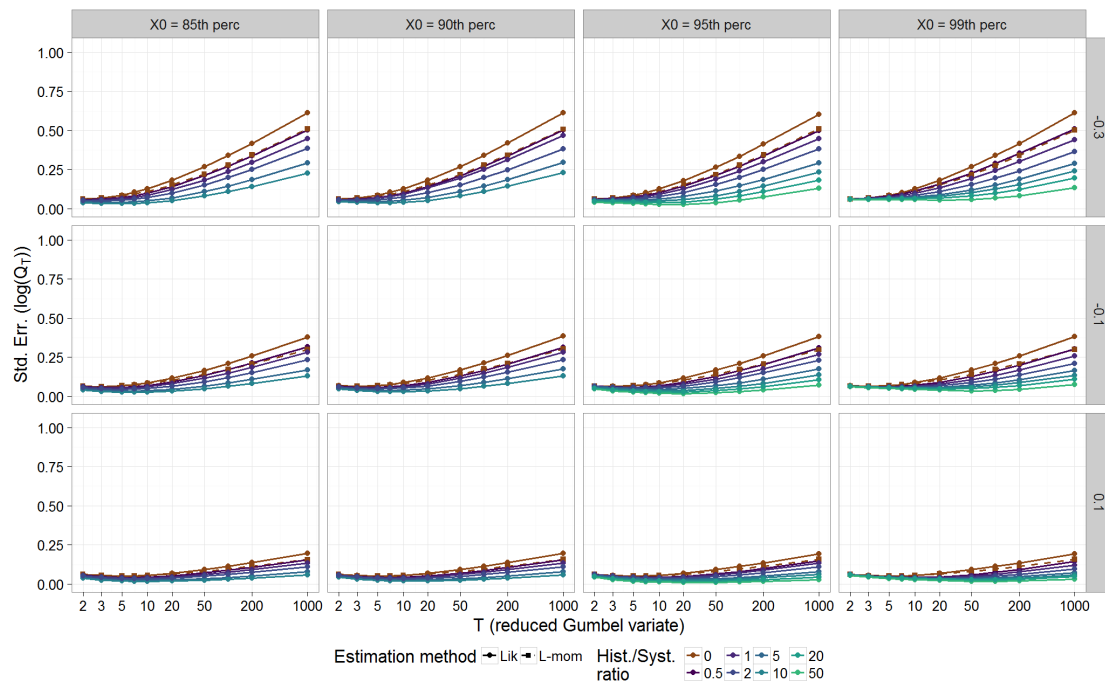
the distribution with non-exceedance probability  $p = 1 - 1/T$ .  $T$  is taken to be a value from among (2, 3, 5, 7, 10, 20, 50, 100, 200, 1000).

The standard errors of the  $\log(Q_T)$  obtained using the maximum likelihood approach are shown in Figures B.10, B.11, B.12 and B.13 as a function of the return period for the different systematic sample sizes. In each plot, the different colours of each line indicate a specific ratio of historic to systematic record length, including the systematic only case for which the ratio is 0. The results obtained with the likelihood approach are shown for all the cases in which systematic and historical data are used; the standard error connected to the at-site estimation using the PWM (L-moments) approach is also shown as a reference. These values are comparable in size to the standard error shown in Kjeldsen (2014b) for estimates of the design floods obtained using the RFFA methods. The return periods on the x-axis are plotted using the reduced Gumbel variate, corresponding to  $-\log(-\log(1-1/T))$ .



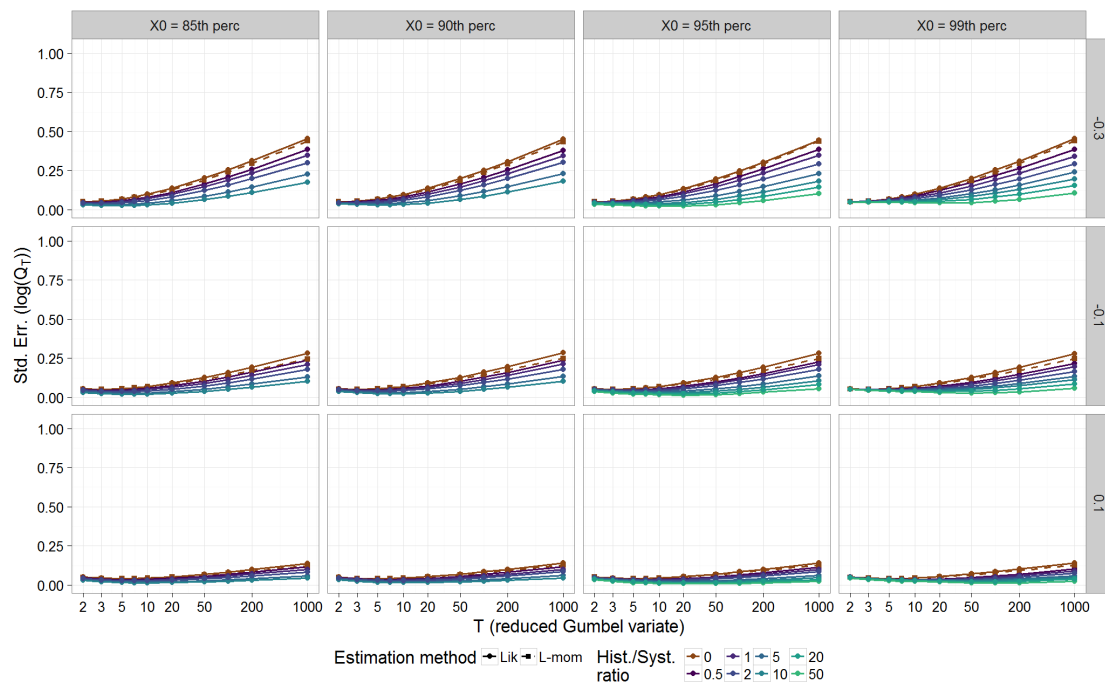
**Figure B.10 Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 10 using the likelihood estimation method**

Notes: Each panel shows different  $X_0$  and shape parameter combination.



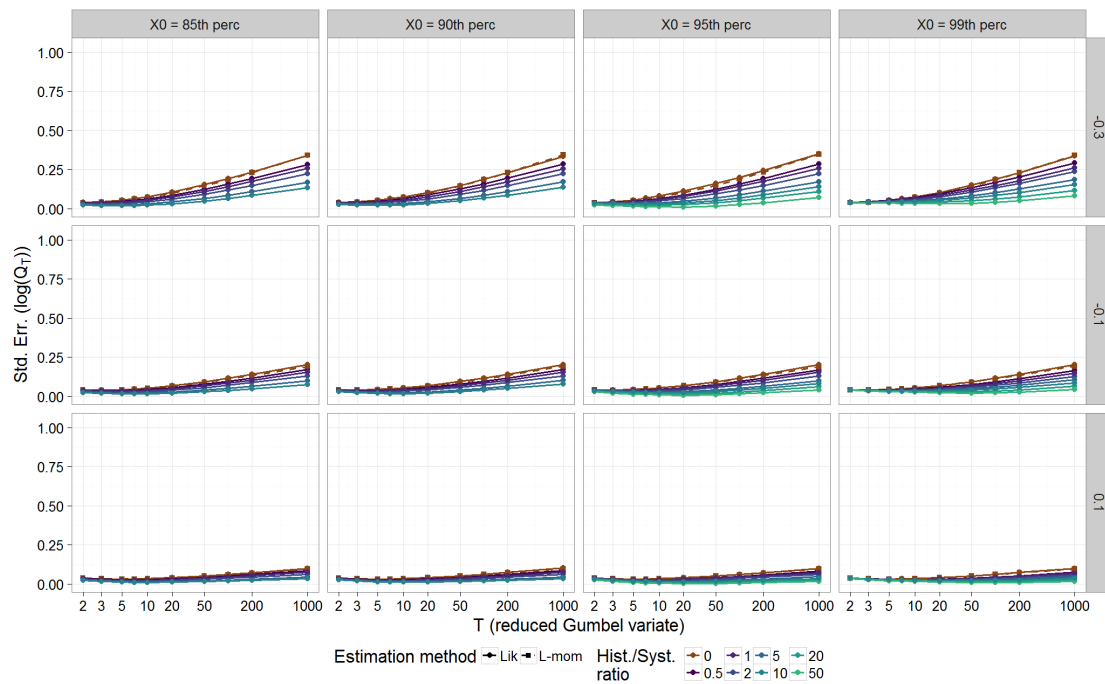
**Figure B.11** Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 30 using the likelihood estimation method

Notes: Each panel shows different  $X_0$  and shape parameter combination.



**Figure B.12** Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 46 using the likelihood estimation method

Notes: Each panel shows different  $X_0$  and shape parameter combination.



**Figure B.13 Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 76 using the likelihood estimation method**

Notes: Each panel shows different  $X_0$  and shape parameter combination.

The uncertainty around all  $Q_T$  values is largely diminished when historical data are included in the estimation procedure in the likelihood framework, for all sample sizes, with the effect being stronger for longer systematic records. If only a short systematic record is available though (10 years, Figure B.10), the use of PWM/L-moments methods on the systematic data gives only a lower standard error, in particular when short, the sample size of the systematic record is small and the historical data are sparse and correspond to exceedances of a high threshold. This very good performance of PWM for short records is the main reason behind the wide use of L-moments in hydrology.

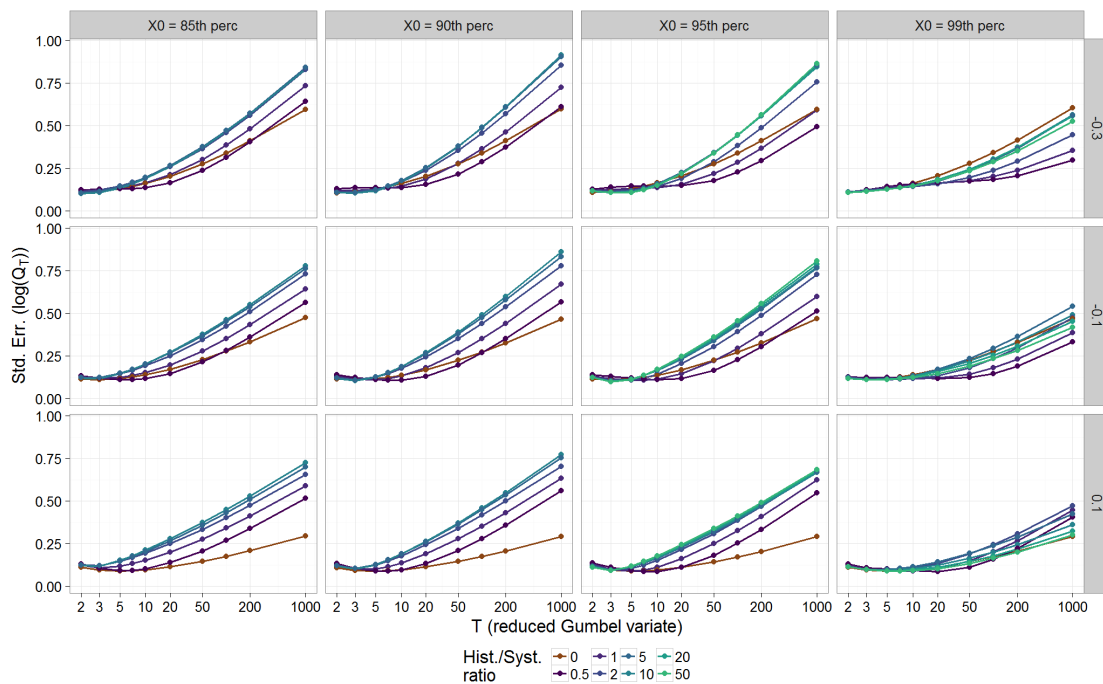
It is also noticeable that errors around the  $Q_T$  tend to be larger for the distribution with a highly negative shape parameter. Since for most peak flow series, the shape parameter is negative, this shows that in many practical cases there is scope to reduce the uncertainty around the estimation of design events with long return periods.

Another interesting property of the profiles of the standard errors is that there seems to be a link between the  $Q_T$  for which the minimum standard error is obtained when using historical data and the return period of the perception threshold. Minimal uncertainty around the  $Q_T$  value is found when the non-exceedance probability  $p = 1 - 1/T$  corresponds to the non-exceedance probability associated to the perception threshold  $X_0$ . Frances et al. (1994) derived the asymptotic variance of  $Q_T$  for the 2-parameter GEV when historical data are included in a maximum likelihood approach and show that some asymptotic relationships exist between  $X_0$ ,  $T$  and the maximum potential gains in terms of uncertainty reduction. A useful indication is that if the main focus of the flood frequency analysis is the estimation of a certain design event,  $Q_T$ , it is likely that historical information on events that exceed the quantile  $Q_T$  would be the most beneficial in terms of uncertainty reduction. Although the real probability of exceedance of the perception threshold cannot be known, it can be roughly estimated by looking at how many exceedance of the threshold have been recorded in the whole sample. Furthermore, the practical implication of this finding is that if a very large design event (that is, a rare event) is the aim of the estimation procedure, it might be beneficial to

focus on few extremely high historical peaks rather than on a more complete sample of average size.

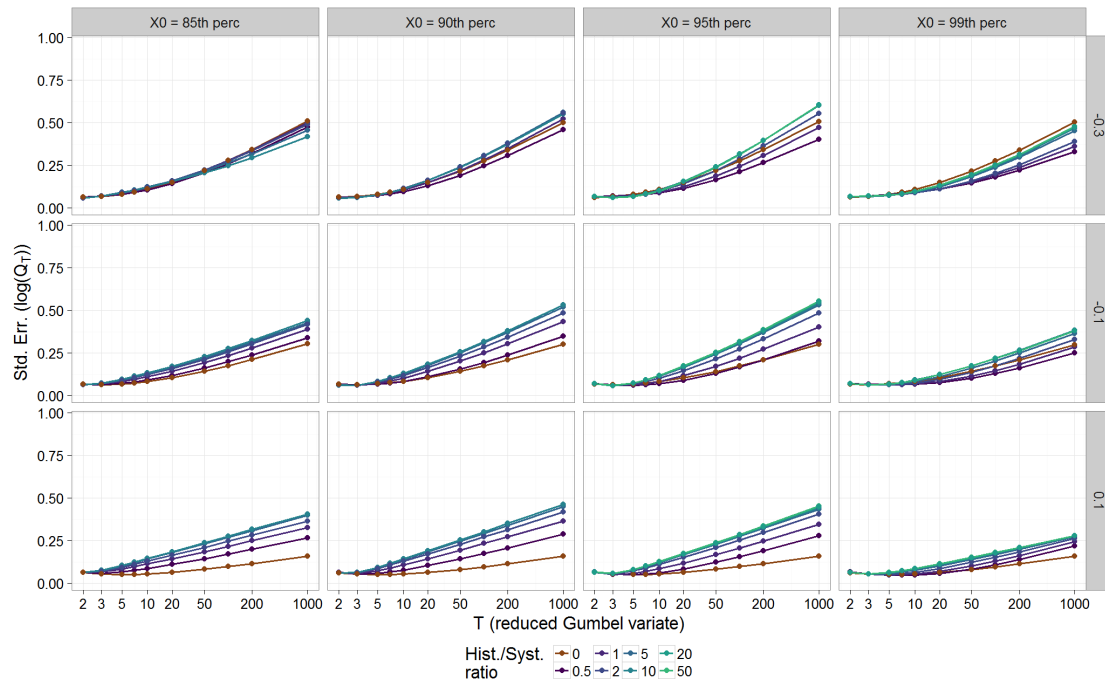
The standard errors of the  $\log(Q_T)$  obtained using the partial probability moment approach are shown in Figures B.14, B.15, B.16 and B.17 as a function of the return period for the different systematic sample sizes. The return periods on the x-axis are plotted using the reduced Gumbel variate, corresponding to  $-\log(-\log(1-1/T))$ .

Using (P)PWM to estimate design events when historical data are available can lead to an increase of the uncertainty around the estimated design event, more so for the case in which the shape parameter is positive, the perception threshold is low and small systematic samples are used. This somewhat surprising result is investigated further in the next section, but overall the results from the simulation study would indicate that the inclusion of historical data might not be beneficial when using (P)PWM. This is in contrast with the original results of Wang (1990a), but it might be related to the fact that (P)PWM methods seem to give a better performance when the data exhibit a higher L-CV than those found in this simulation study.



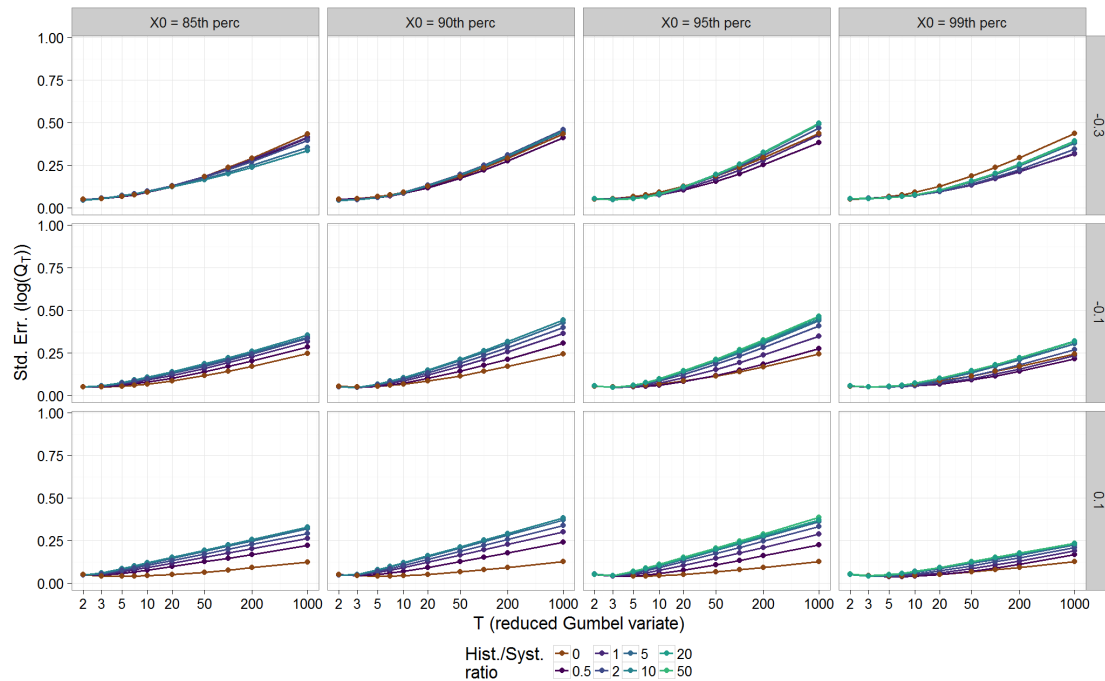
**Figure B.14** Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 10 using the (P)PWM estimation method

Notes: Each panel shows different  $X_0$  and shape parameter combination.



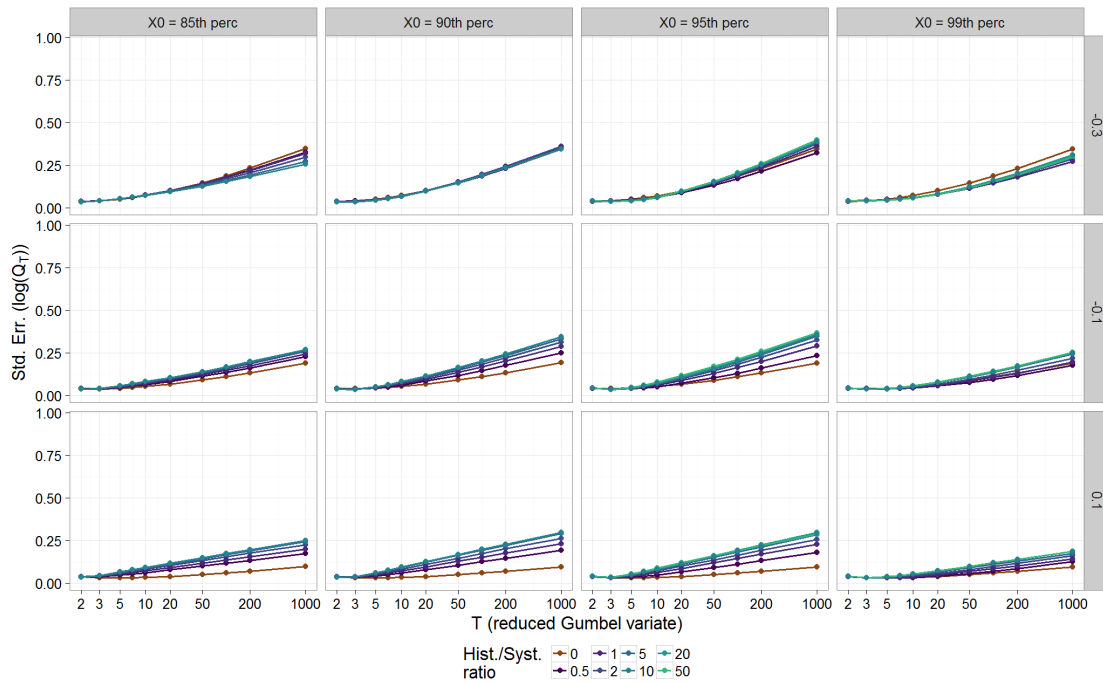
**Figure B.15** Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of using the (P)PWM estimation method

Notes: Each panel shows different  $X_0$  and shape parameter combination.



**Figure B.16** Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 46 using the (P)PWM estimation method

Notes: Each panel shows different  $X_0$  and shape parameter combination.



**Figure B.17 Standard error for the  $\log(Q_T)$  as a function of the return period for a systematic sample size of 76 using the (P)PWM estimation**

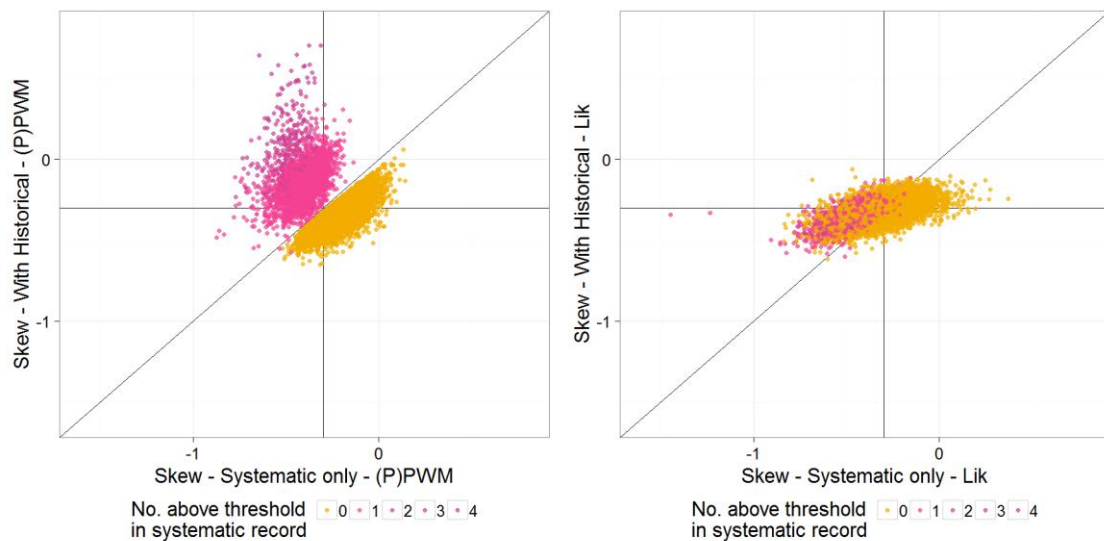
Notes: Each panel shows different  $X_0$  and shape parameter combination.

## B.6 Effect of high events in the systematic record

One interesting aspect of the estimation results is not visible when showing the summary results, but becomes evident when the  $S = 10,000$  shape parameter estimates obtained with historical exceedances are plotted against the corresponding estimates obtained using only the systematic data as in Figure B.18. Each data point in the figure is coloured according to the number of points in the systematic record which exceed the perception threshold ( $X_0$ , which corresponds to 100-year event in this case).

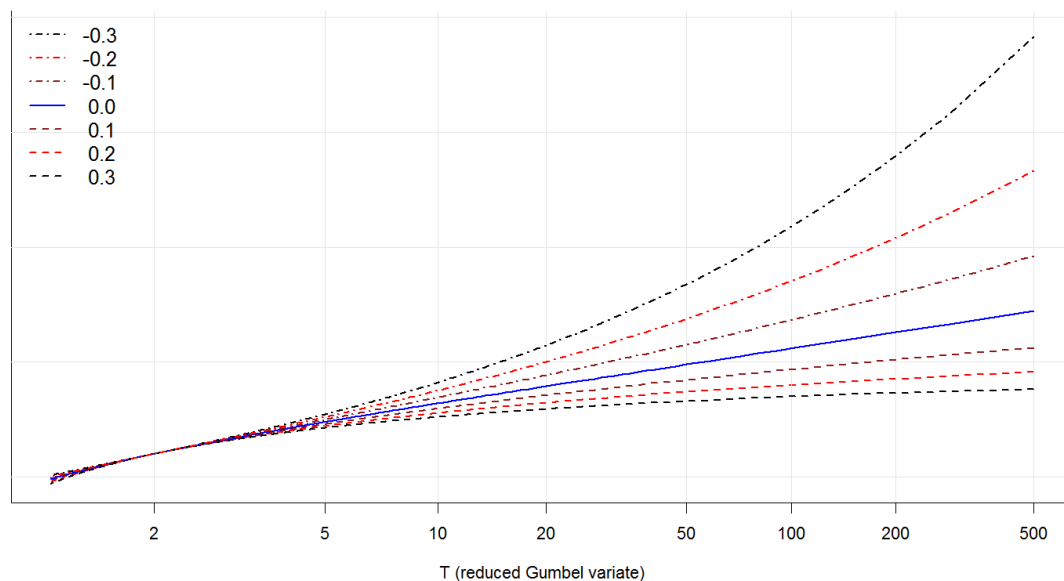
Two clusters can be distinguished in the left panel of Figure B.18 (which corresponds to the (P)PWM approach): if the systematic data under study have no exceedance of the perception threshold, the inclusion of some historical data will result in smaller estimated (more negative) values of the shape parameter, that is, steeper curves (see Figure B.19). In contrast, when at least one exceedance of the perception threshold was recorded in the systematic data, larger estimates of the shape parameter are obtained when historical data are included in the analysis.

This clustering is not as evident in the likelihood-based results (Figure B.18, right panel), although a similar pattern can be observed. The reasons behind this evident clustering lie in the mathematical formulation of the (P)PWM; however, it is an important feature of the data analysis. Indeed, if no exceedance of the perception threshold is recorded, the estimation procedure would tend to estimate a value for the shape parameter that is higher than the one under which data were generated (that is, estimating a flatter flood frequency curve); the available data do not justify a very small (for example, large absolute value) shape parameter estimate. The presence of at least one threshold exceedance in the data from the historical records has the definite effect of changing the shape parameter estimate.



**Figure B.18 Estimated shape parameter when using systematic data only or systematic and historical data. Left: estimates obtained with (P)PWM method. Right: estimates obtained with Likelihood Method**

Notes: Systematic record length: 30 years  
Historic record length: 300  
 $X_0$  = 99th percentile



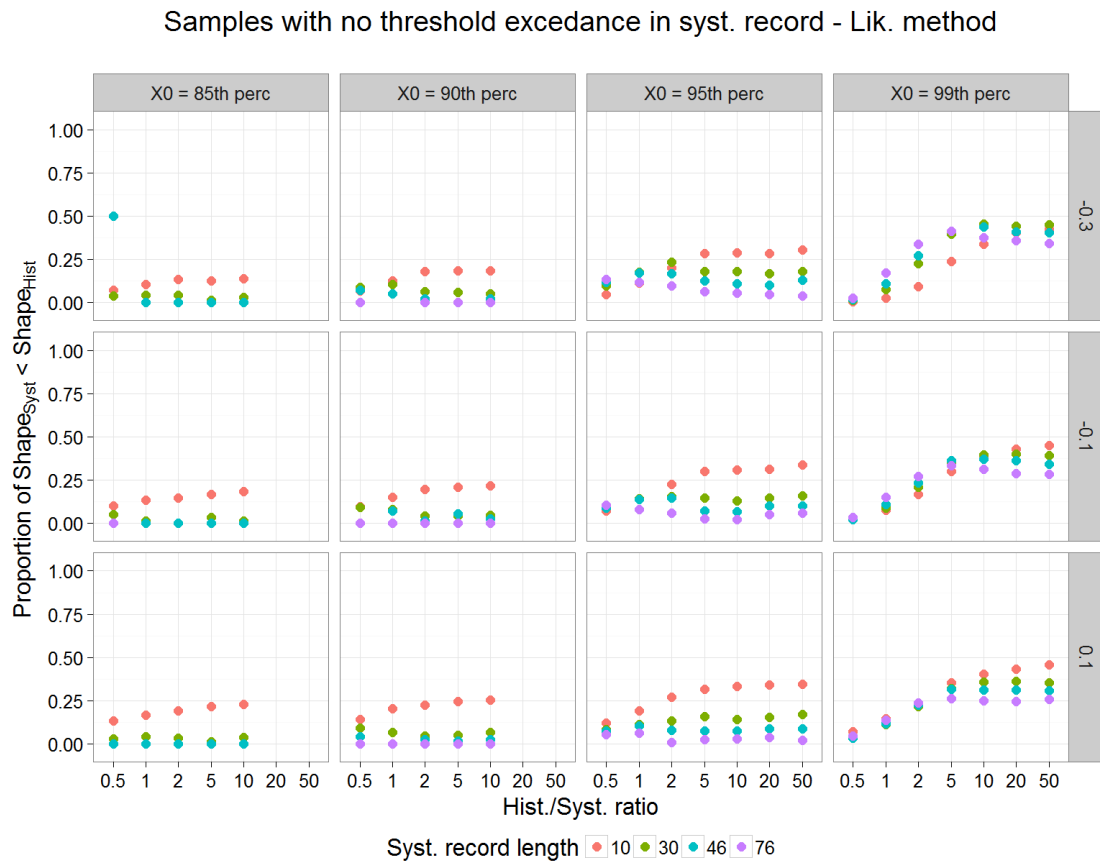
**Figure B.19 Relationship between the shape parameter and the flood frequency curve (GLO distribution)**

For all simulation settings when using the (P)PWM estimation procedure, the shape parameter estimate is always smaller when historical data are used to augment a systematic record with no exceedance of the perception threshold. This is not the case for the likelihood estimation procedure, although there is a tendency in the same direction (see Figure B.20). For a large proportion of cases in which one or more exceedance of the perception threshold was recorded in the systematic data, however, using some historical data in the estimation procedure results in a higher estimate for the shape parameter (that is, flatter flood frequency curves, see Figure B.19) when using either the likelihood or the (P)PWM estimation procedure (see Figure B.21 and Figure B.22, respectively).



For both cases, the proportion of cases in which the use of historical data lead to higher estimate of a shape parameter is increasing for larger perception thresholds with long historical records. Note that the dots in the plots only indicate proportions within the samples for which one or more systematic samples include some observation higher than the perception threshold. According to the sample size and the threshold, there will be a smaller or larger number of cases in which the threshold is exceeded. In some instances for the simulated series of 10 years of data, no sample had any observations higher than the perception threshold when this corresponded to the 99th percentile.

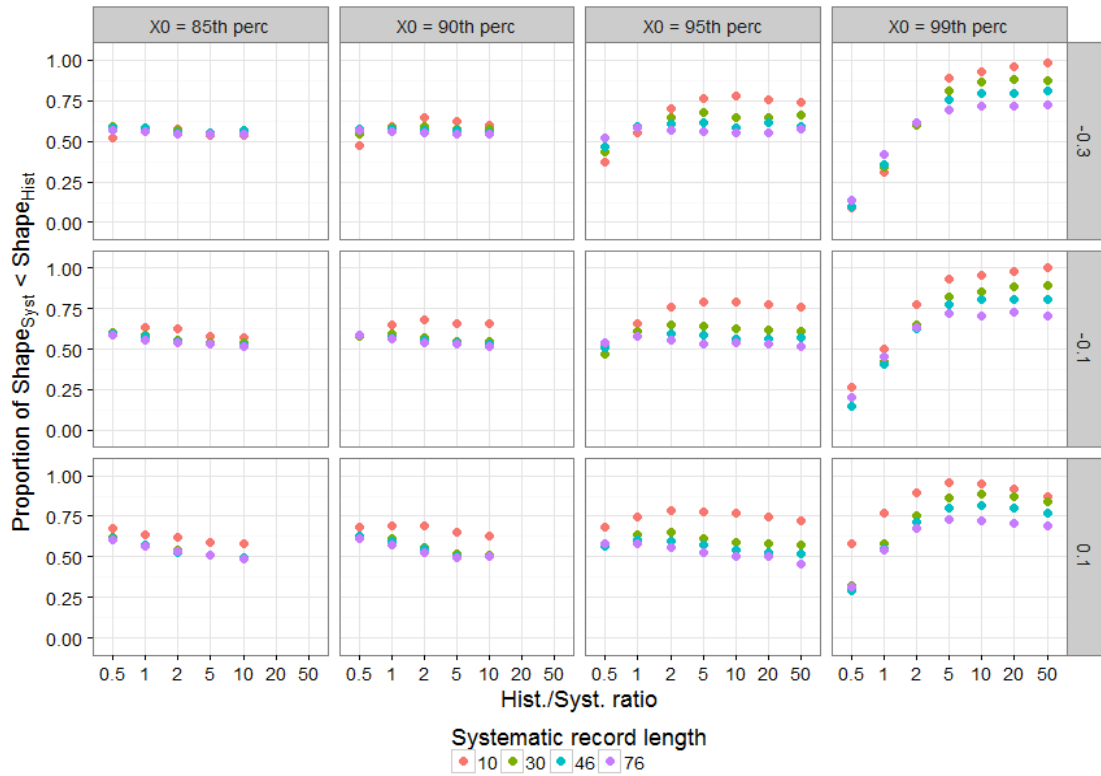
The overall properties of the performance of the estimation procedures are affected by a number of factors. However, it is interesting to note that some properties of the data at hand can already give an indication of how adding information on historical data will affect the final estimate. If the historical data available correspond to a flooding event that is bigger than any observed event in the record, the natural effect would be for the estimated flood curve to become steeper, that is, design events would become bigger. But if the events available from historical information correspond to events of a size not bigger than what is present in the systematic record, the effect of using the additional information can produce either steeper or flatter curves depending on a series of properties of the data.



**Figure B.20** Proportion of cases in which the estimate of the shape parameter using systematic and historical data is larger than when using only systematic data in simulated records for which no event in the systematic record exceeded the perception threshold  $X_0$  using the likelihood estimation method

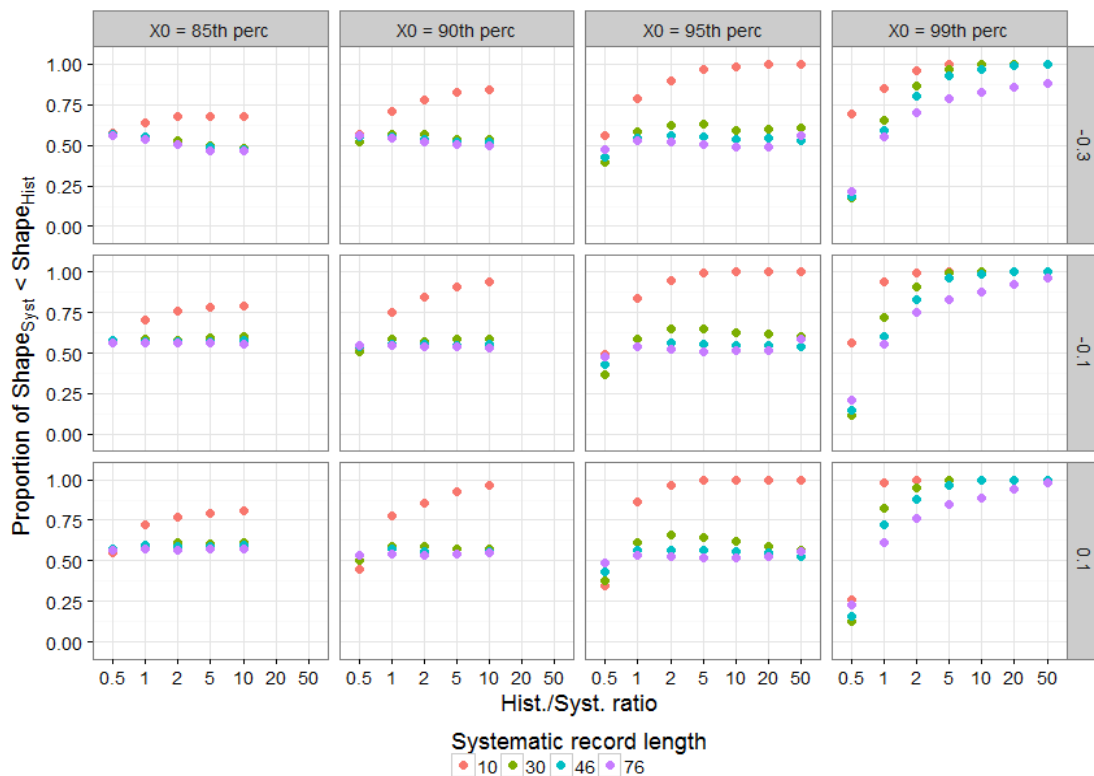


# Samples with at least one threshold exceedance in syst. record - Lik. method



**Figure B.21** Proportion of cases in which the estimate of the shape parameter using systematic and historical data is larger than when using only systematic data in simulated records for which at least one event in the systematic record exceeded the perception threshold  $X_0$  using the likelihood estimation method

### Samples with at least one threshold exceedance in syst. record - (P)PWM method



**Figure B.22** Proportion of cases in which the estimate of the shape parameter using systematic and historical data is larger than when using only systematic data in simulated records for which at least one event in the systematic record exceeded the perception threshold  $X_0$  using the (P)PWM estimation method

## B.7 Sensitivity analysis

In flood frequency estimation procedures, a number of pieces of information need to be included in the model and a number of assumptions are made on the basis that all this information is correct. Namely, it is assumed that:

- the information available covers a whole period  $h$
- all the  $k$  exceedances above the perception threshold  $X_0$  are known, possibly with a peak flow value attached to each threshold exceedance

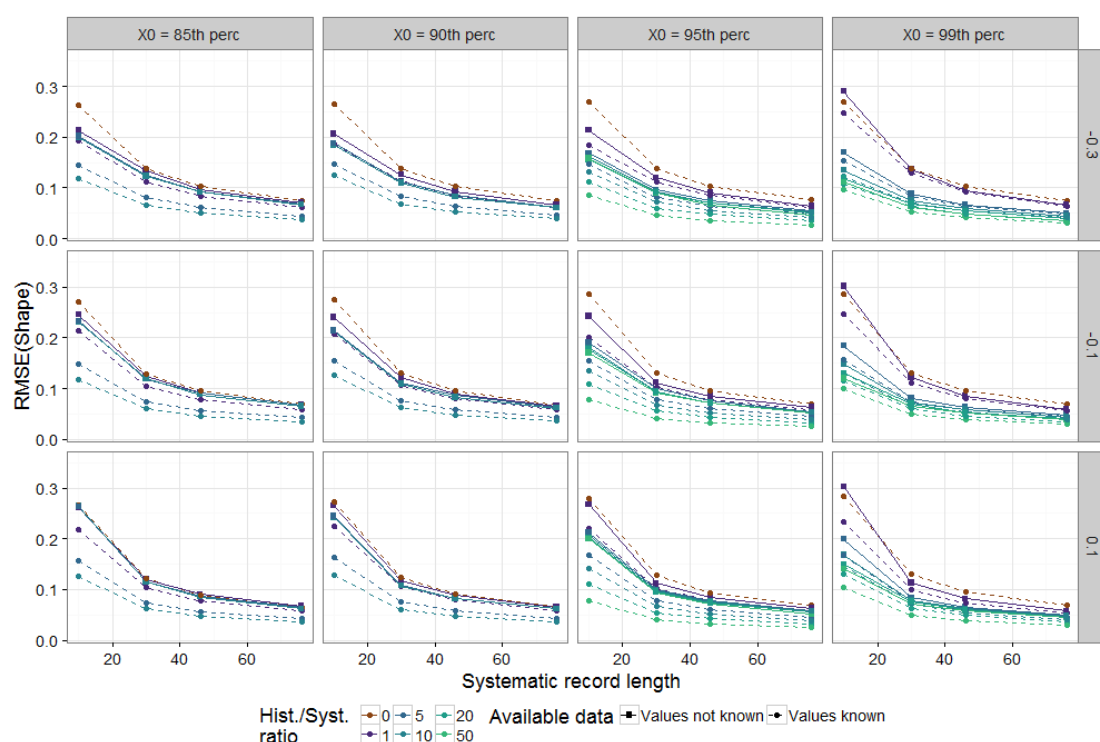
The simulation results presented above show that, when all the information needed in the model estimation is correctly known, overall there seems to be a reduction in the uncertainty around the estimates of design events, in particular if maximum likelihood methods are employed. This section examines the sensitivity of these improvements against possible mis-specifications of some of the quantities involved in the estimation procedure.

Each of the datasets simulated under the previous setting are re-analysed using some different choice of a specific quantity in the estimation procedure. In some situations, the settings with systematic record of length 76 were dropped to make the computations feasible.

## B.7.1 Sensitivity analysis – threshold exceedances and binomial censoring

As mentioned previously, the magnitude of historical floods often cannot be reliably estimated from the information available and all that is known is that some threshold was exceeded. The information on the past floods is therefore only partially complete and is censored in the sense that only information on the magnitude that was exceeded is available. Partial probability weighted moments cannot be easily modified to include information of this type, while the likelihood approach can be readily adjusted to account for this type of information.

Some investigation of the effect of knowing or not knowing the value of the historical floods is given in Figure B.23, where the RMSE for the shape parameter estimated when the historical values are known or not known are shown as a function of the systematic record length. To make the figure slightly less cluttered, the results for some ratios of historical to systematic length are dropped.



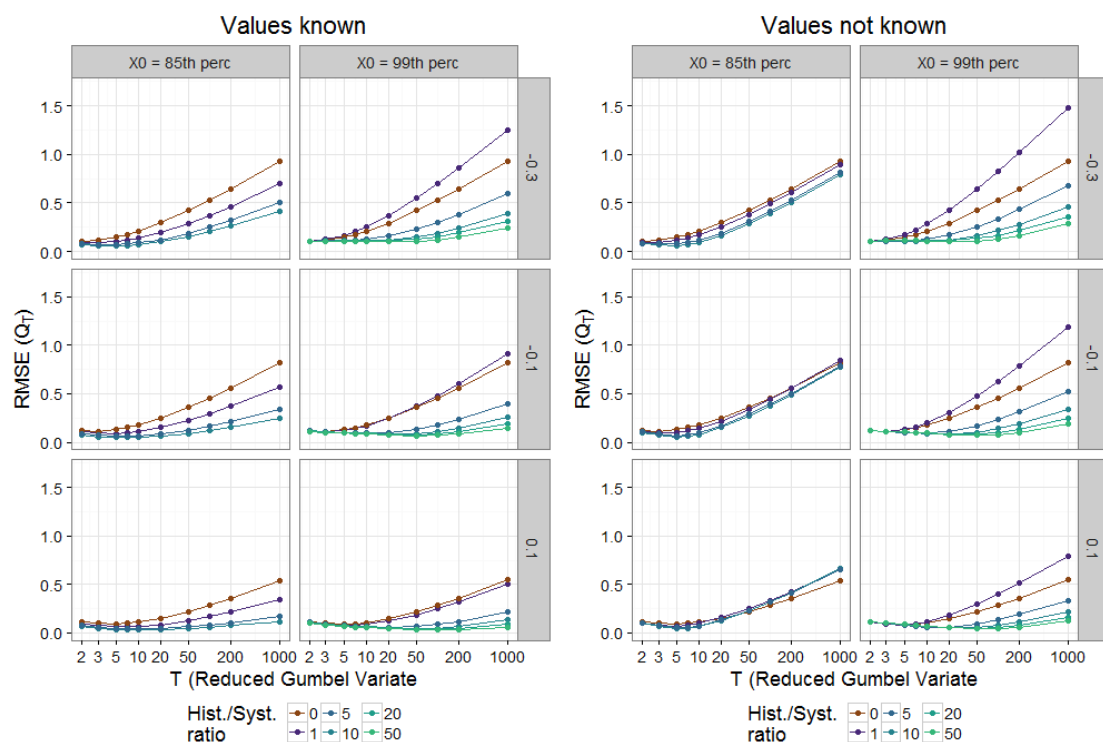
**Figure B.23 RMSE for the shape parameter when using known values of the historical data or when only the number of perception threshold exceedances is known as a function of the systematic sample size**

Notes: Estimation method: likelihood

Knowledge of the value of the high flows of the historical record seems to be valuable information when a low perception threshold is used, with much smaller RMSE values for the case in which the flow magnitudes are known. When only very rare past events are recorded, the performance of the estimation method is only slightly improved by the knowledge of the flow magnitudes, in particular when longer systematic records are available. Some differences in performance are still visible even for high perception thresholds when very long historical records are available. This corresponds to the case in which a fairly large number of threshold exceedances are recorded in the simulated data. Essentially, the effect of knowing the actual value of the flow magnitude is bigger for cases in which a fairly large number of events above the threshold are recorded.

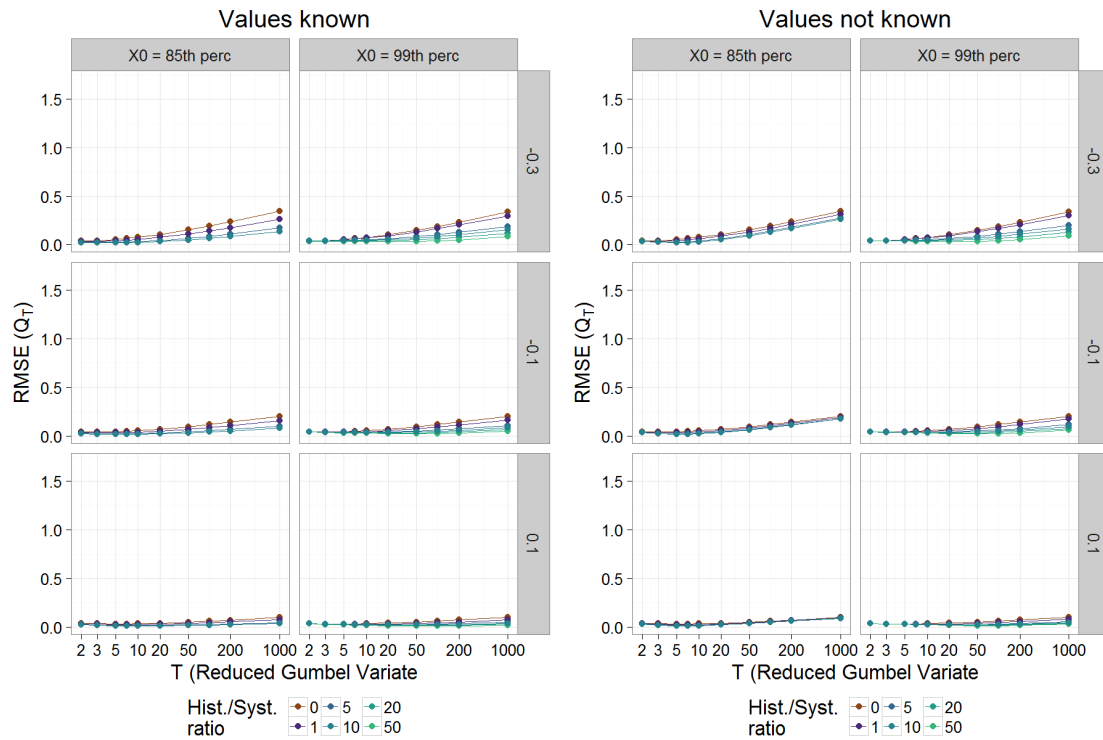
A similar conclusion can be drawn by looking at the RMSE for the  $\log(Q_T)$  obtained in the case in which the values are known or not known for systematic records of 10 and 76 years with a selected perception threshold (Figure B.24 and Figure B.25, respectively). Knowing the values of the threshold exceedances helps in reducing the uncertainty around events with long return periods, in particular if a large number of events above the threshold are available. The effect is less marked for short return periods across all thresholds.

One important aspect of the availability of the information of historical flow values that has not been investigated here is the potential impact of errors in the flow value estimation. Given that some uncertainties exist around the effective value of events from the past, it is possible that including some inexact information of the magnitude of the historical flow could eventually lead to less precise estimates. When the number of threshold exceedances is small and only a high perception threshold is exceeded, it is likely that simply adding the information on the number of exceedances would give similar reduction in uncertainties as knowing the peak flow values.



**Figure B.24 RMSE for the  $\log(Q_T)$  with a systematic sample size of 10 when using the known values of the historical data or when only the number of perception threshold exceedances is known**

Notes: Estimation method: likelihood



**Figure B.25 RMSE for the  $\log(Q_T)$  with a systematic sample size of 76 when using the known values of the historical data or when only the number of perception threshold exceedances is known**

Notes: Estimation method: likelihood

## B.7.2 Sensitivity analysis – incomplete information

Although research for information on past floods might seek to obtain a complete census of past flooding events, some events may not get picked up in the historical record. This section investigates the possible effects of including incomplete historical records in the estimation procedure.

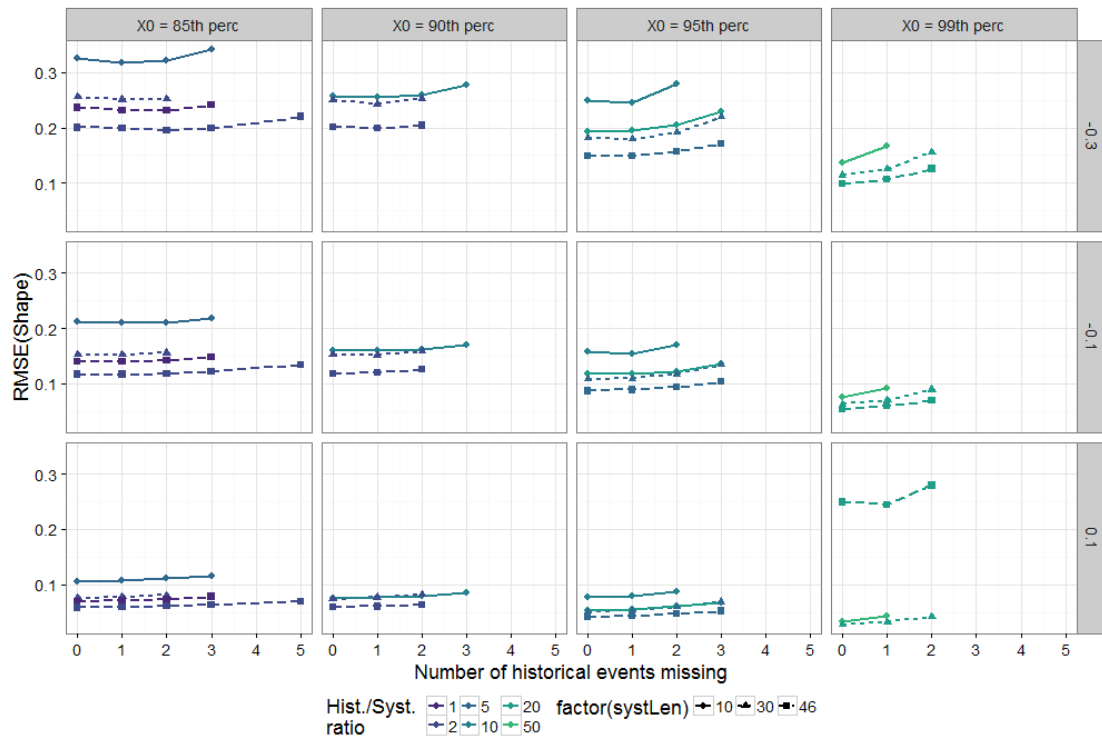
The conceptualisation of the testing framework within the simulation study for this exercise is not trivial, as a random number  $k$  of threshold exceedances is recorded in the Monte Carlo generation of each historical sample. The simulation setting forced at least one historical record above the threshold to be generated, but it did not fix the number of exceedances of the perception threshold.

The expected number of threshold exceedances in a given historical record of length,  $h$ , with a perception threshold corresponding to the  $p^{\text{th}}$  percentile is  $E[K] = h * (1 - p)$ . The sensitivity of the estimation procedure to the presence of incomplete information (for example, historical records not included in the estimation procedure) was evaluated only on the simulation settings for which the expected value of events above the threshold is larger than 5 ( $k > 5$ ). Furthermore, for each simulation setting the number of events that could potentially have not been included in the historical record was different: if the total number of threshold exceedances is 5, it is not possible to generate a dataset in which more than 4 historical record are missing and still have some historical information.

Depending on the expected number of events in the historical record, different sets of events of increasing size  $k^-$  were deleted from the historical record, with  $k^-$  varying within (1, 2, 3, 5, 6, 10, 15, 30, 45). For each simulation setting, the values that  $k^-$  spans vary depending on the expected number of threshold exceedances. The effect of

deleting the information for  $k^-$  events on the final estimate depends on the actual true value of  $k$ .

A subsample of results for the simulation settings with low expected number of threshold exceedances  $k < 15$  is shown in Figure B.26, where the RMSE for the shape parameter is shown as a function of the number of events deleted from the historical record. Only selected simulation settings are shown in each panel and, for each simulation setting, the expected number of historical events to exceed the threshold is different. Each line type and symbol indicates a different sample size, while colours indicate the historical to systematic length ratio. For example, in the upper left panel, the upper blue line with dots indicates the simulation setting in which a systematic record of length 10 years is augmented by an historical record covering a period of 50 years, in which one would expect to observe on average  $(1 - 0.85) \times 50 = 7.5$  events above the threshold. The greenish line with dots just below indicates the simulation setting in which a systematic record of length 10 years is augmented with an historical record covering a period of 100 years, in which one would expect to observe on average  $(1 - 0.85) \times 100 = 15$  events above the threshold. Thus it is possible to test the effect of excluding up to 6 events from the historical record. The effect of not including one or more historical data points can be assessed by looking at the increase in the RMSE as the number of missing data points increases.

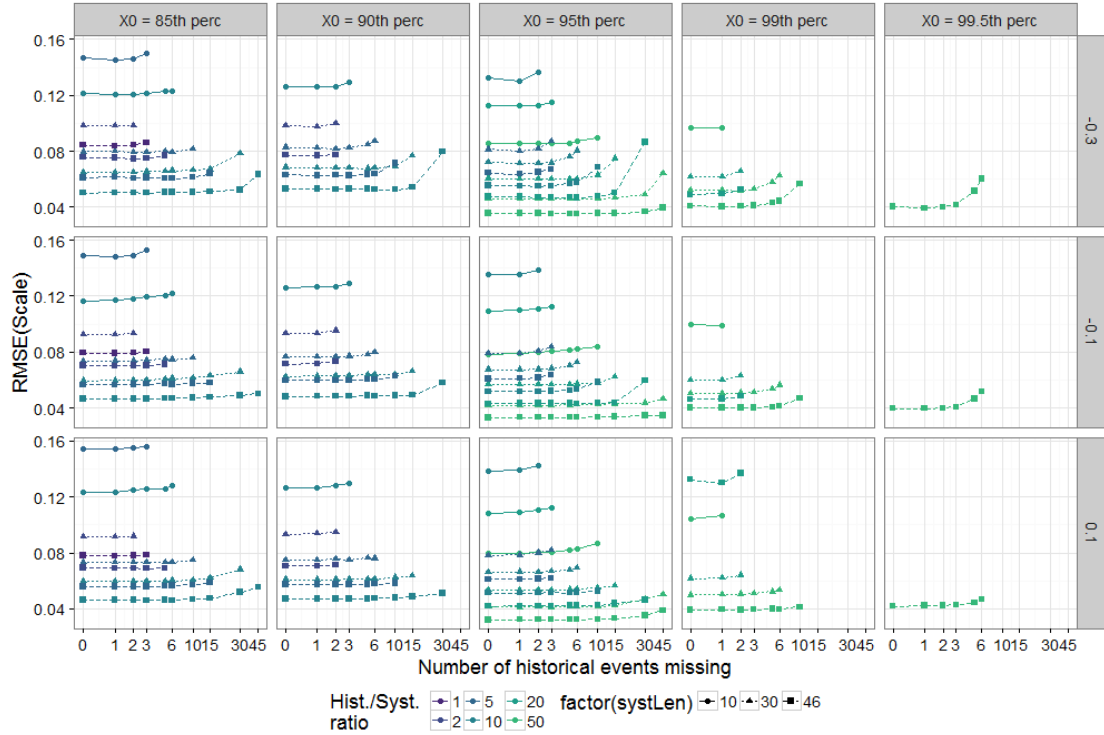


**Figure B.26 RMSE for the shape parameter as a function of the number of historical events missing in the historical record using the likelihood estimation method**

Notes: Original historical records contain on average at most 15 points.

The effect of missing information in the historical sample does not seem to have a very large impact on the quality of the estimate of the shape parameter, although in the examples shown in Figure B.26, in most cases the number of events discarded in the historical sample tend to be a relatively small proportion of the whole historical information. However, since considerable effort is taken when constructing series of historical events, it is hoped that only a small percentage of the past events are not present in the dataset.

To give a more complete description of the relationship between the number of missing data points and the loss in terms of the RMSE of the shape parameter, the results for a larger set of simulation settings is given in Figure B.27. The x-axis is tweaked into a log scale to make the figure slightly more readable. Note that some additional simulation settings with higher perception threshold  $X_0$  were included to investigate the effect of the incomplete samples; the expected number of threshold exceedances for these simulation setting is not very high for very long historical periods, corresponding to more realistic data availability scenarios.

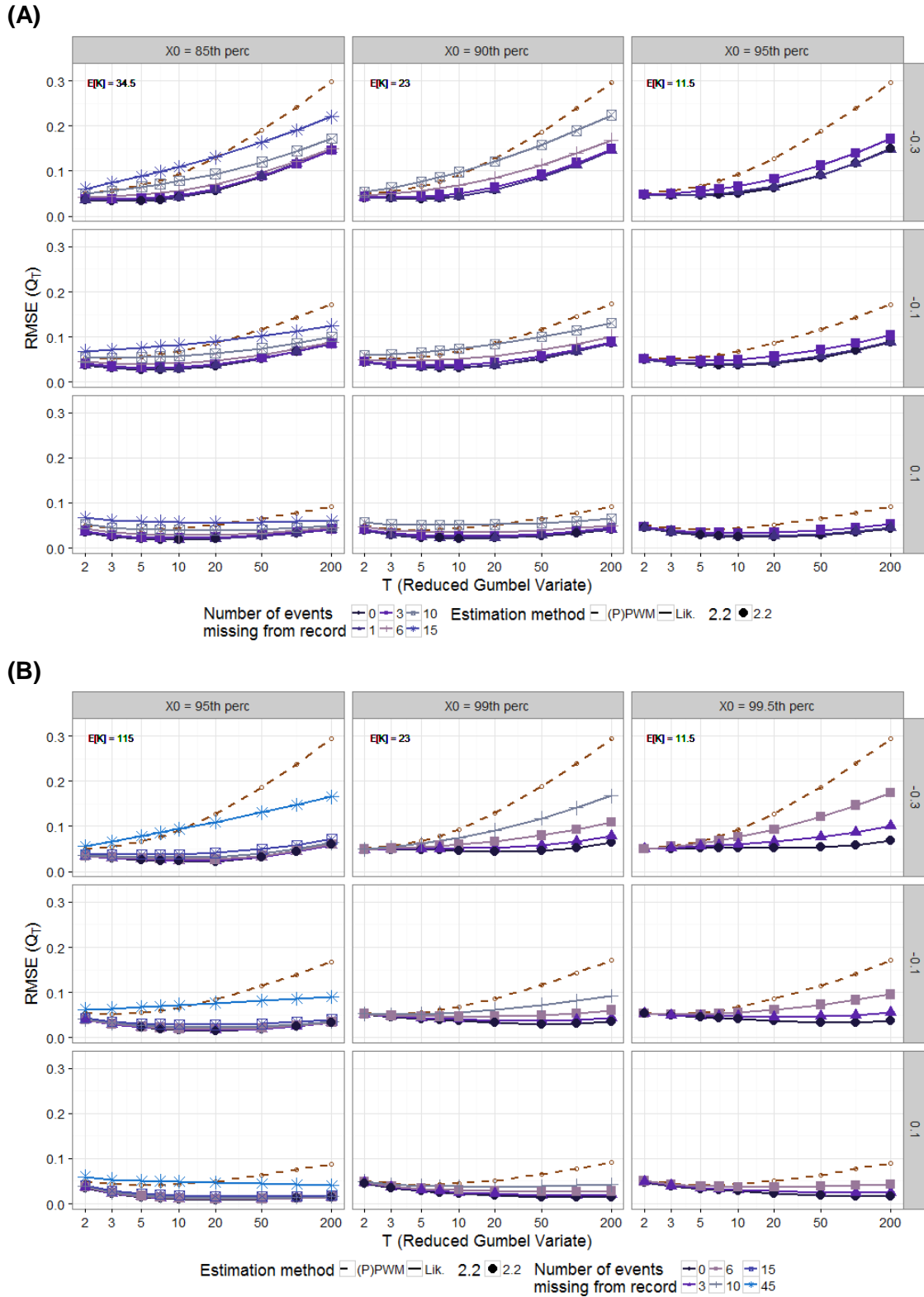


**Figure B.27 RMSE for the shape parameter as a function of the historical events missing in the historical record using the likelihood estimation method**

The estimation procedure appears to be relatively resistant to the case in which a fairly high number of historical events is missing from the historical record. The estimation gives much larger RMSE only when very large proportions of the historical events are missing. For example, in the upper right panel of Figure B.27, the long dashed line with squares represents the results for the case in which a systematic record of 46 years is augmented with an historical record spanning  $h = 2,300$  years. Since the perception threshold corresponds to the 100-year event, on average there would be 23 historical events exceeding  $X_0$ . The estimation performs in a fairly stable way for increasing number of missing data points up to  $k^- = 6$  and only gives visibly worse results when 10 points are missing from the historical record. This corresponds to almost half of the historical sample. The effect of the missing information is stronger in the case of negative shape parameters.

To investigate the overall performance of the likelihood method using historical data on incomplete samples for design event estimation, the RMSE for the log of the estimated of different  $Q_T$  values obtained when using a systematic sample of 46 years and an historical sample of 230 years and 2,300 samples is shown in Figure B.28. The RMSE obtained when using the PWM/L-moment method with the systematic record only is also displayed for reference.





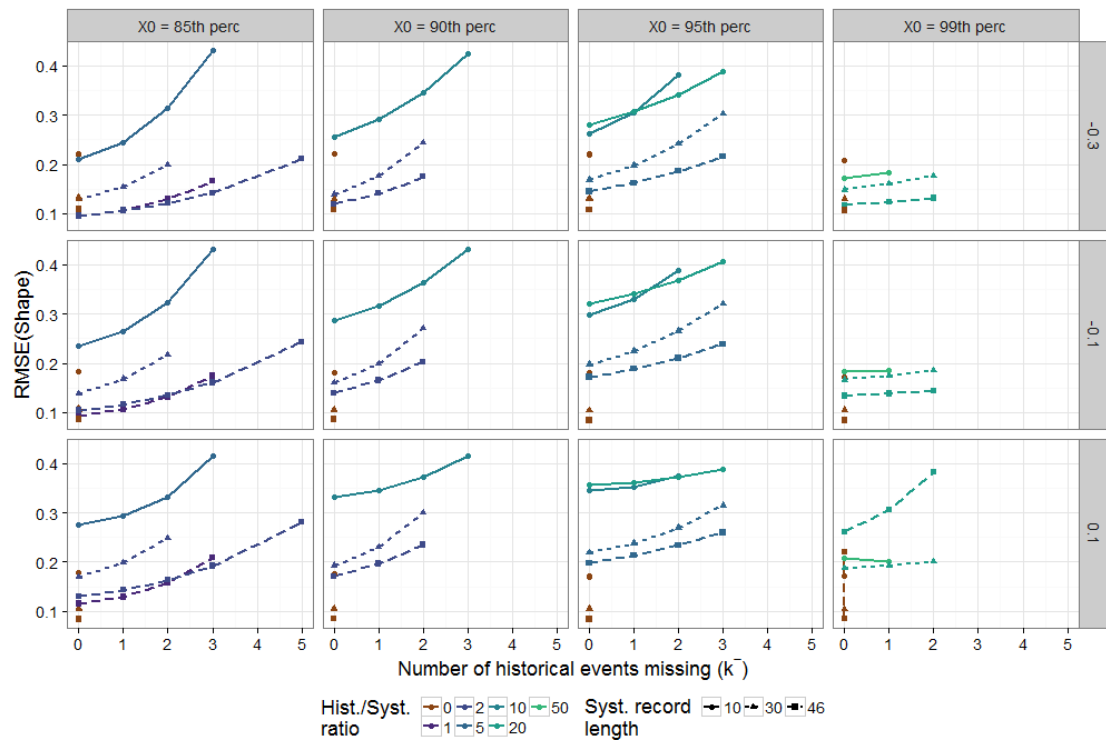
**Figure B.28 RMSE for the  $\log(Q_T)$  for a systematic sample size of 46 and an historical period covered of (A)  $h = 230$  and (B)  $h = 2,300$**

Notes: The expected number of historical events is shown for each perception threshold  $X_0$ . Colours indicate the number of events not included in the historical record  $k^-$ . The dashed line indicates the PWM estimate for systematic data only. Estimation methods: likelihood.



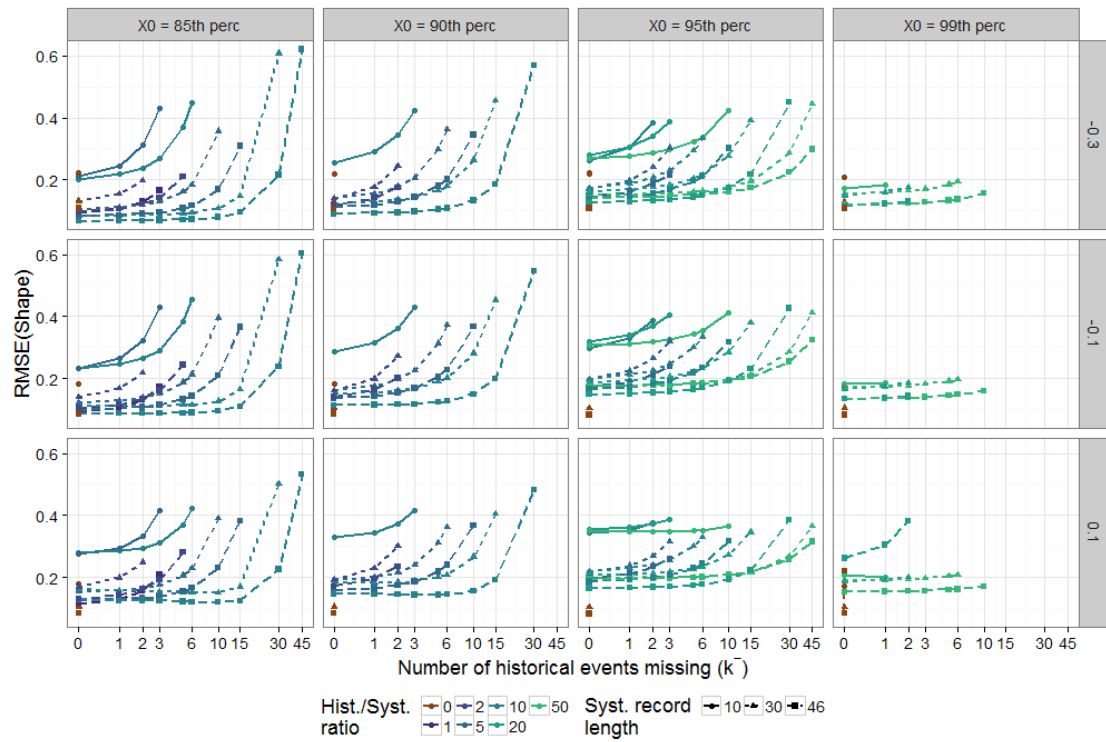
The effect on the overall performance of the lack of threshold exceedances in the historical record is stronger for lower perception thresholds and for distributions with negative shape parameters. Interestingly, when events with high return period are to be estimated the estimation seems to give lower RMSE than the traditional L-moment analysis even if a large part of the historical sample is missing.

Results for the effect of the lack of historical information in the (P)PWM setting are shown in Figure B.29 and Figure B.30, which show the effect of the increasing number of missing data points in the historical record on the RMSE for the shape parameter in the (P)PWM approach. These figures can be compared with Figure B.26 and Figure B.27. The (P)PWM method seems to be less robust than the likelihood method to the lack of information in the historical record.



**Figure B.29 RMSE for the shape parameter as a function of the historical events missing in the historical record using the (P)PWM estimation method**

Notes: Original historical records contain on average at most 15 points.



**Figure B.30 RMSE for the shape parameter as a function of the historical events missing in the historical record using the (P)PWM estimation method**

### B.7.3 Sensitivity analysis – uncertain historical record period

When retrieving information on past events which affected the catchment under study, every effort should be made to gather a complete view on the realistic coverage of the historical information available. Nevertheless, it is possible that no clear beginning of the length of time period covered by the historical series  $h$  is available. In cases where no information on  $h$  is available, it is possible to estimate the length of the historical period  $h$  from the properties of the threshold exceedances.

In the case of only one threshold exceedance in the whole historical period, Strupczewski et al. (2014) suggested estimating its value taking  $\hat{h} = 2 * t$ , where  $t$  indicates the length of the period between the time of the historical record and the beginning of the systematic series. The motivation behind the proposed estimate stems from the idea that the time at which the perception threshold is exceeded is uniformly distributed on a discrete domain  $[0, 1, \dots, h]$ . Since the expected value of a uniform distribution on  $[0, h]$  ( $H \sim U(0, h)$ ) is equal to  $E[H] = (h - 0)/2$ , when only one observation is available the suggested estimate for  $h$  is  $2 * t$ .

It can be shown that this estimate corresponds to using a constrained L-moment estimate for the upper limit of a uniform distribution. Hosking and Wallis (1997) gave the relationship between the first 2 L-moments ( $\lambda_1, \lambda_2$ ) and the parameters of a uniform distribution on the continuous domain  $[\alpha, \beta]$ ,  $U(\alpha, \beta)$  as:

$$\lambda_1 = \frac{1}{2}(\alpha + \beta) \quad (\text{equation B.21})$$

$$\lambda_2 = \frac{1}{6}(\beta - \alpha) \quad (\text{equation B.22})$$

from which the following relationships can be derived:

$$\beta = 2\lambda_1 - \alpha \quad (\text{equation B.23})$$

$$(\beta - \alpha) = 6\lambda_2 \quad (\text{equation B.24})$$

The practical case of estimating the starting year of the historical record could be reframed as the estimation of the value of  $\beta$  after having fixed  $\alpha$  to 0, since the beginning of the gauged record is known.

If the sample of timing of threshold exceedance is composed of only one record of value  $t$ , only the first sample L-moment  $l_1$  can be computed and it corresponds to  $l_1 = t$ . The estimate for  $\beta$  can then be taken to be  $\hat{\beta} = 2l_1 = 2t$ , which corresponds to the estimate proposed in Strupczewski et al. (2014).

If more than one threshold exceedance is recorded in the historical period, the distance between the beginning of the systematic record and the timing of the historical exceedances is recorded in the sample  $(t_{(1)}, \dots, t_{(n)})$ , where samples are ordered in decreasing order from the largest value of  $t$  (that is, the first historical record) to the smallest (that is, the most recent historical record). From the sample of time records, sample values for the first 2 L-moments  $l_1$  and  $l_2$  can be obtained. Keeping  $\alpha$  fixed at 0, an estimate for  $\beta$  can be again be found as  $\hat{\beta} = 2l_1$ . This estimate is here called L1 estimate.

However, rather than fixing the  $\alpha$  parameter, an estimate for the interval width  $w = (\beta - \alpha)$  could be taken to be  $\hat{w} = 6l_2$ . The width  $w$  can then be plugged in as an estimate for the historical record length  $h$ . This estimate is here called L2 estimate.

Both estimation approaches suffer from the drawback that there is no formal assurance that the estimated value of  $h$  is actually larger than  $t_{(1)}$ , the time of record of the first historical event. This can be resolved in practice by taking the estimate of  $h$  to be the maximum between  $t_{(1)}$  and the preferred estimate of  $\hat{h}$ . Furthermore, if more than one historical record is available, the estimated value of  $h$  would belong to the continuous scale rather than the discrete scale, which can be easily fixed in practice by rounding the  $\hat{\beta}$  value.

Since the sample  $l_1$  value corresponds to the average value of a sample, the L1 estimate corresponds to twice the average distance between the time of record of the historical values and the beginning of the systematic sample. This value is relatively easy to compute and easier to communicate than the L2 estimate. Furthermore, the L2 estimate can only be calculated if at least 2 threshold exceedances are present in the historical sample. The performance of the L2 estimate has been investigated in the simulation study, but little difference from the performance of the L1 estimate was found and is not discussed further.

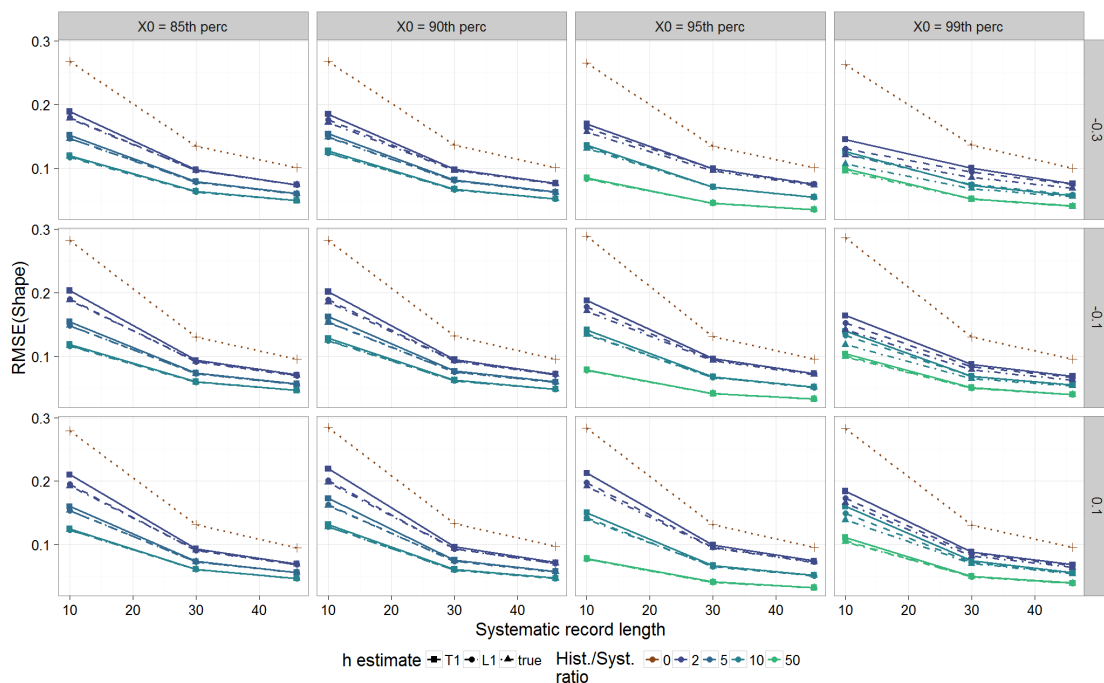
Once again, next to the use of L-moments, another possible approach to the statistical estimation of a parameter characterising a distribution is maximum likelihood. The maximum likelihood estimate for the left boundary of a uniform distribution corresponds to the minimum of the sample. In the practical application at hand, this corresponds to taking  $h = t_{(1)}$ , or, in other words, having the period covered by historical information to start at the time of recording of the first historical record. Here this estimate is called the T1 estimate. The use of the T1 estimate is widely discouraged in the literature on the inclusion of historical data for flood frequency analysis, but it is important to acknowledge that a statistical motivation for such estimate could be given, especially in the unlikely case of a large number of historical events.

All the approaches to the estimation of  $h$  presented in this section (L1, L2 and T1 estimates) rely on the assumption that the timing of the threshold exceedances is uniformly distributed. This is a realistic assumption when the process describing the flood magnitude is stationary, that is, it is equally likely at any point in time to record a threshold exceedance. Deviations from this assumption – due to either the natural alternation between flood rich and flood poor periods, or a higher likelihood of not

recording events which happened further away in time – would undermine the performance of all estimation approaches. The use of a statistical estimation for the historical period length should only be employed when it is truly impossible to identify a sensible point in time at which it is credible that all events above a threshold are present in the historical sample.

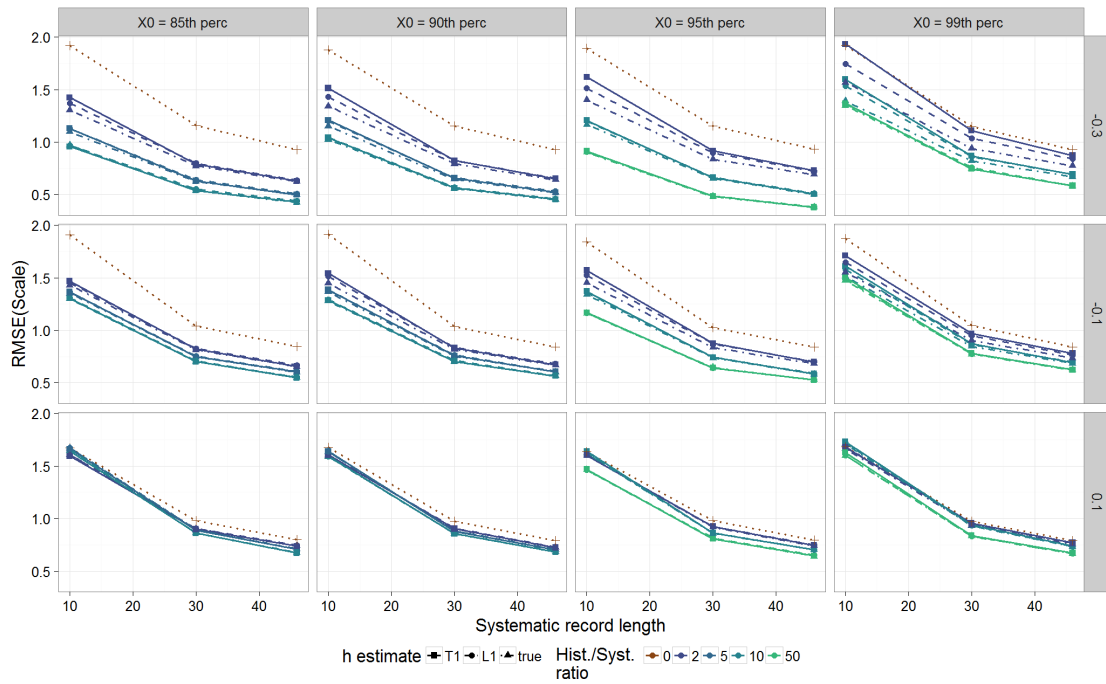
Figure B.31 shows the RMSE for the estimated shape parameter and Figure B.32 the estimated scale parameter as a function of the systematic record length for selected lengths of the historical records when using different approaches to estimate the value of  $h$  within the likelihood framework. The RMSE values obtained when using the true value of  $h$  and when using systematic data only are also shown.

The bigger differences in the performance for the different approaches to the estimation  $h$  can be seen for shorter records and higher perception threshold. The historical samples in these cases is fairly small and all approaches are likely to give a poor estimate of  $h$ , the T1 approach in particular. Once the historical sample size is larger, however, using an estimated value of  $h$  gives a similar performance to using the true value of  $h$ , with little difference between the T1 and L1 estimation.



**Figure B.31 RMSE for the shape parameter as a function of the systematic record length for selected historical record lengths**

Notes: Line types and shapes indicate the estimation approach used to estimate  $h$ . Each panel shows a different  $X_0$  and shape parameter combination. Estimation method: Likelihood.



**Figure B.32 RMSE for the scale parameter as a function of the systematic record length for selected historical record lengths**

Notes: Line types and shapes indicate the estimation approach used to estimate  $h$ . Each panel shows a different  $X_0$  and shape parameter combination. Estimation method: Likelihood.

#### B.7.4 Sensitivity analysis – incorrect perception threshold

While gathering information about the historical floods, it is likely that some information can be obtained on what is the minimal flood magnitude above which floods are likely to have left some traces. This value corresponds to the perception threshold,  $X_0$ , and in the standard procedures, it is assumed that all events above the threshold which occurred in the historical period are known.

An initial investigation of the effect of using a wrongly identified perception threshold is given below, though once again it is challenging to conceptualise simulation settings that can give a fair representation of the issues which might arise in practical cases.

For each synthetic dataset with historical data generated in the simulation study, new estimates for the model parameters are obtained using a perception threshold,  $X_0^l$ , which is different from the correct one. The true  $X_0$  values were defined as specific quantiles of the data generating distribution, namely the 85th, 90th, 95th and 99th percentiles. Three types of modification are made to the perception threshold:

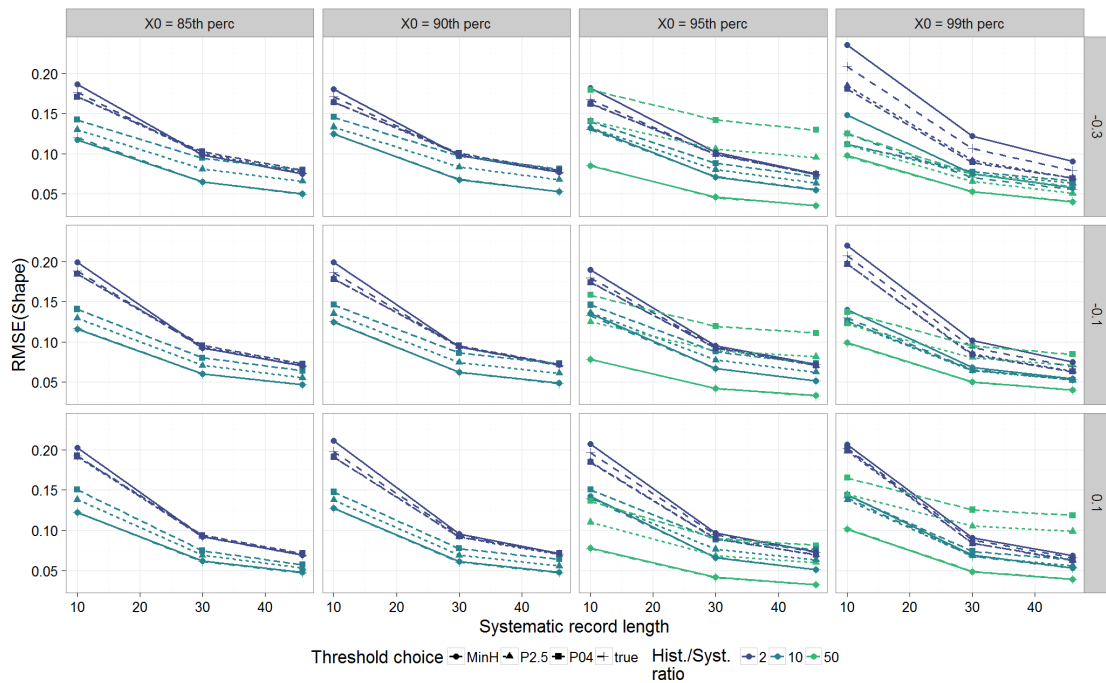
- the threshold used in the estimation corresponds to a value just below the lowest flood in the historical record
- the perception threshold is taken to be a value corresponding to a percentile 2.5% lower than the true one
- the perception threshold is taken to be a value corresponding to a percentile 4% lower than the true one

For example, in a situation in which the true perception threshold in a simulation with shape parameter equal to  $-0.3$  was taken to be the 85th percentile, the true  $X_0$  value

corresponds to 47.23. However, the estimation carried out using a perception threshold  $X_0^l$  corresponding to the 82.5th percentile and the 81st percentile gave values of 45.35 and 44.36, respectively. The 3 different modification of the perception threshold are called here MinH, P2.5 and P04. By construction, the threshold actually used in the estimation under the MinH option would be different for each simulated sample.

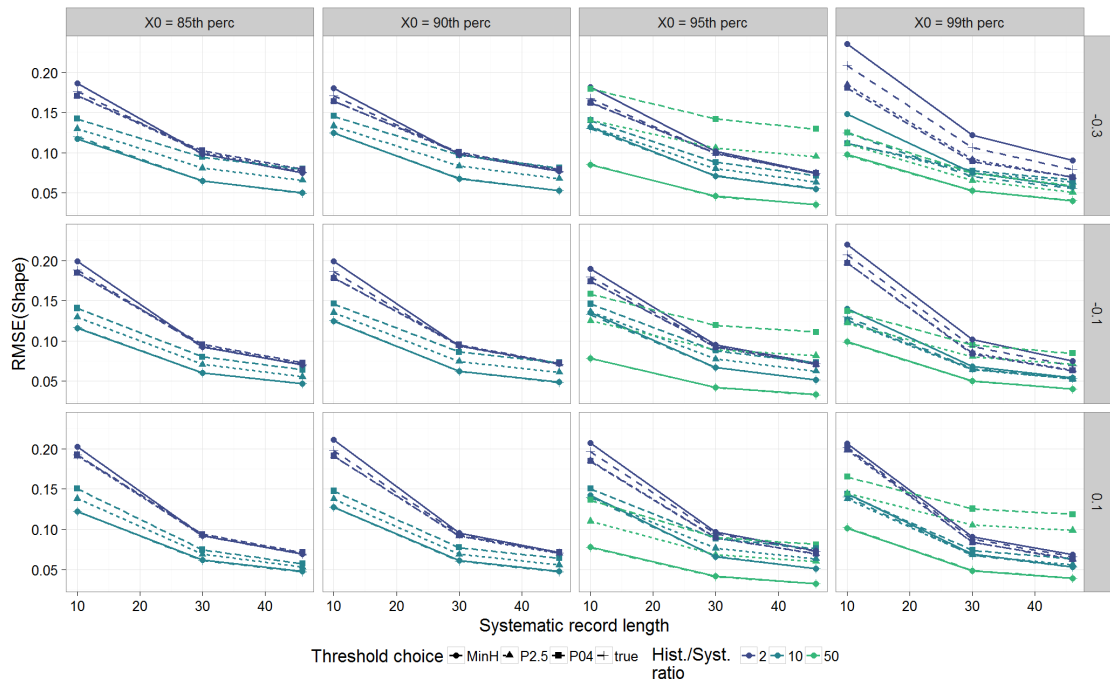
Figures B.33 and B.34 show the RMSE and the Bias, respectively, for the estimated shape parameter as a function of the systematic record length for selected lengths of the historical records when different strategies are used to select the perception threshold used in the estimation procedure.

The results are quite confused, but overall there seem to be quite an important effect on the quality of the estimation of the shape parameter when the threshold used in the estimation is not the correct one. The impact is larger when longer historical periods are available. For long historical records, the use of the MinH strategy can deliver a low value of RMSE, but this seems to be less true when the historical record covers only a short period. For shorter historical periods, especially when the real perception threshold is high (that is, few historical data points are available), the P2.5 and P04 strategies actually give lower RMSE values than when using the true  $X_0$  value.



**Figure B.33 RMSE for the shape parameter as a function of the systematic record length, for selected historical record lengths**

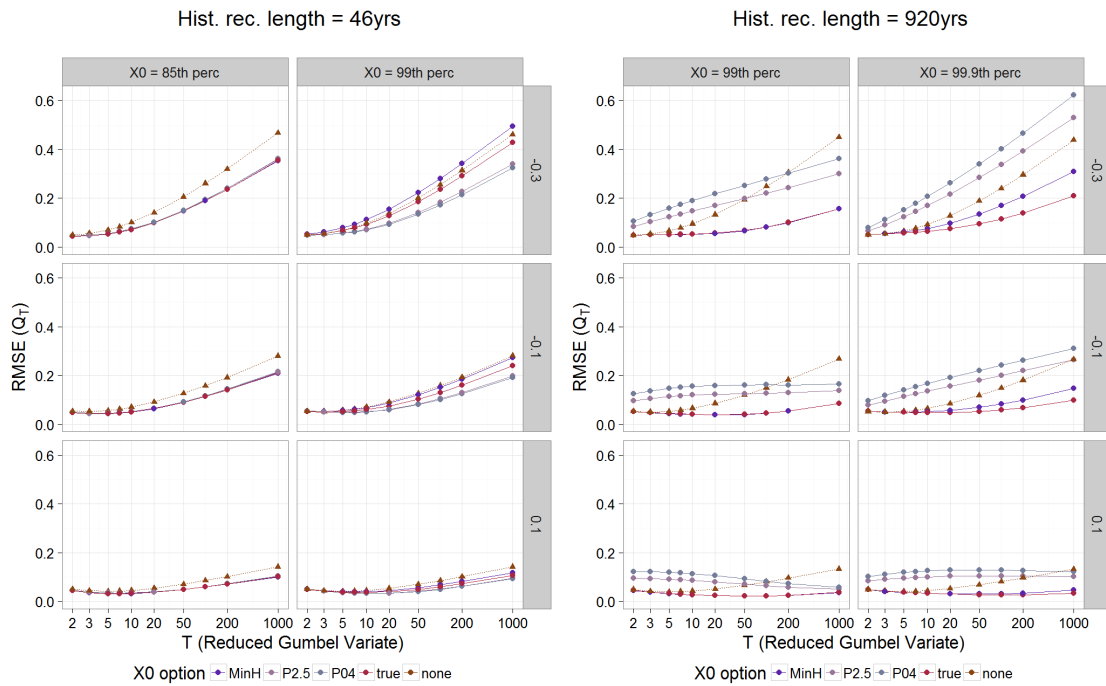
Notes: Line types and shapes indicate the estimation approach used to estimate  $h$ .  
Each panel shows a different  $X_0$  and shape parameter combination.  
Estimation method: Likelihood.



**Figure B.34 Bias for the shape parameter as a function of the systematic record length for selected historical record lengths**

Notes: Line types and shapes indicate the estimation approach used to estimate  $h$ .  
Each panel shows a different  $X_0$  and shape parameter combination.  
Estimation method: Likelihood.

The overall effect of the choice of perception threshold on the performance of the estimation of design events is illustrated in Figure B.35, where the RMSE for the  $Q_T$  values with the systematic sample record equal to 46 and historical record covering 46 years (left panel) and 920 years (right panel) are shown. The values obtained when using PWM/L-moments on the systematic data only are also shown for reference.

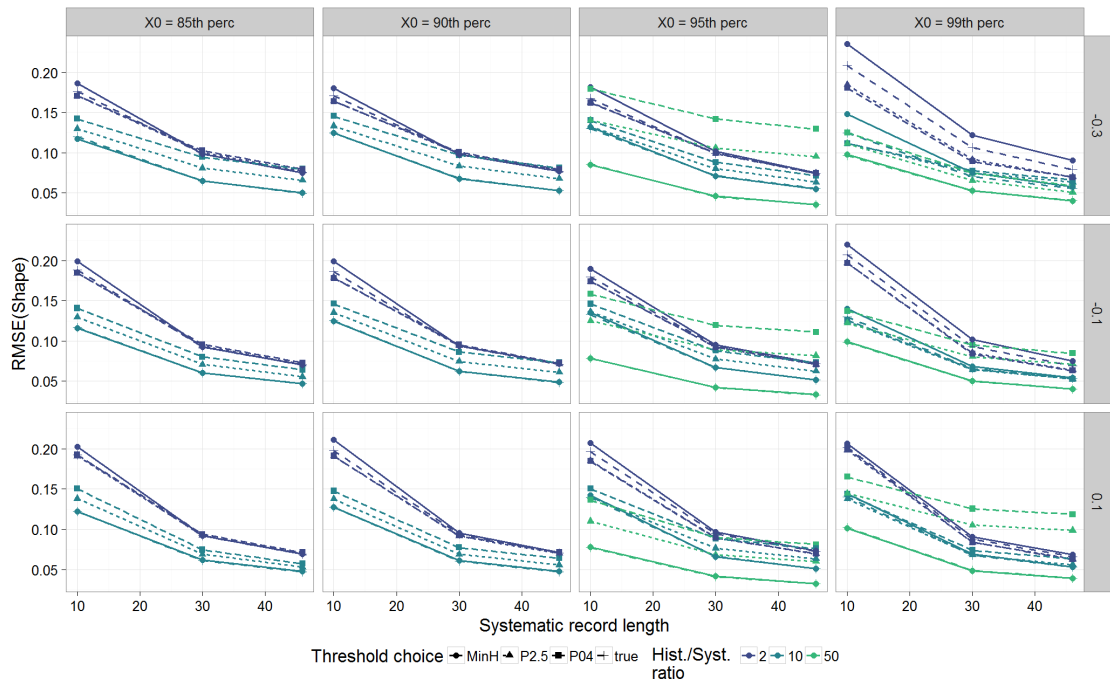


**Figure B.35 RMSE for the  $\log(Q_T)$  for a systematic sample size of 46 and an historical period covered of  $h = 46$  (left panel) and 920 (right panel)**

Notes: Colours indicate the option used to select the perception threshold actually used in the estimation.  
The dashed line indicates the PWM estimate for systematic data only.  
Estimation method: Likelihood.

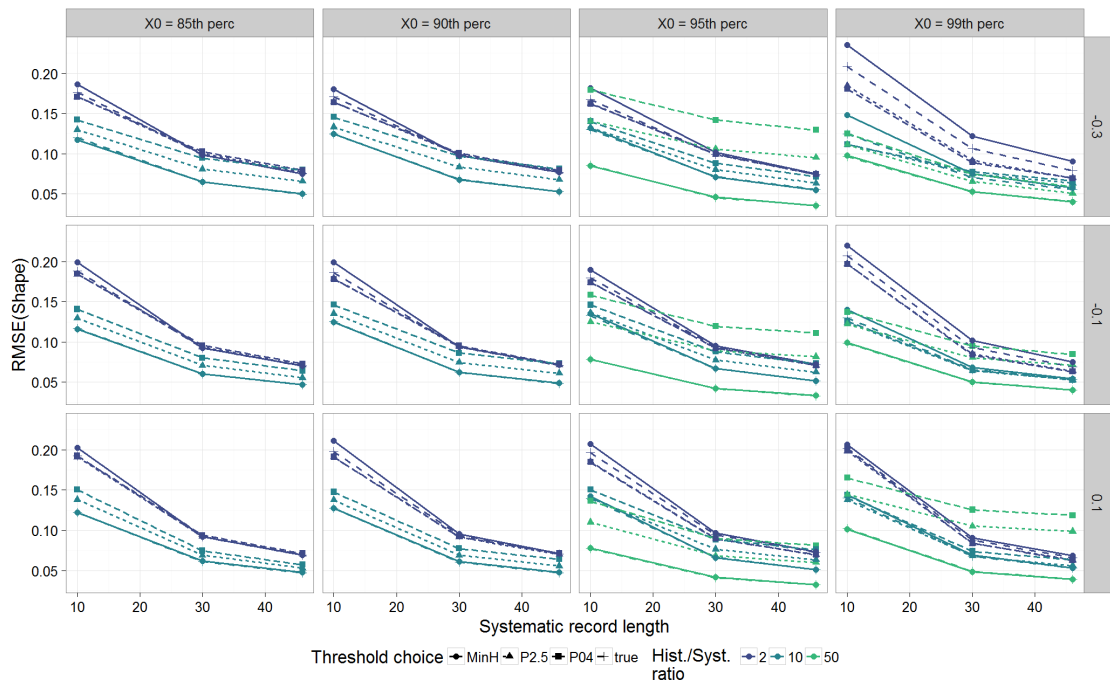
Overall, it seems that using an incorrect perception threshold can have a major impact on the quality of the estimate and can, in some cases, undermine the utility of integrating historical data in the in flood frequency analysis. This is even more true when the (P)PWM approach is used, as it can be seen from Figures B.36 and B.37.





**Figure B.36 RMSE for the shape parameter as a function of the systematic record length for selected historical record lengths**

Notes: Line types and shapes indicate the estimation approach used to estimate  $h$ . Each panel shows a different  $X_0$  and shape parameter combination. Estimation method: Likelihood.



**Figure B.37 Bias for the shape parameter as a function of the systematic record length, for selected historical record length**

Notes: Line types and shapes indicate the estimation approach used to estimate  $h$ . Each panel shows a different  $X_0$  and shape parameter combination. Estimation method: Likelihood.

## B.8 Historical data and pooled analysis: initial considerations

A natural question that might arise is how the estimate obtained by augmenting systematic data with historical information compares with that obtained using the routinely used FEH statistical pooled method. The aim of both approaches is to augment the data available at a site of interest with some additional information which should aid the development of more precise flood frequency estimates. However, comparing the uncertainties connected with each method is not straightforward. The utility of historical data will depend very much on the length of the historical period and the number of past peak events. It is therefore difficult to give a unique indication of the potential gains of a more widespread use of historical data in flood frequency estimation.

Much of the work on the inclusion of historical data in flood frequency analysis focuses on at-site estimation. It is not yet clear how to combine estimates obtained by augmenting systematic records with historical information via a maximum likelihood with the standard FEH methods which rely on L-moments and the pooling of different stations. In Environment Agency (2008), the estimated distribution parameters are obtained from the known relationships between distribution parameters and sample L-moments. The L-moments used in the estimation are obtained as weighted averages of the at-site L-moments of each station of the pooling group. With the maximum likelihood approach, the distribution parameters are estimated directly; however, these could be translated into equivalent sample L-moments via the known relationships.

An initial possible avenue to combine estimates from the maximum likelihood within a pooling approach would be to simply calculate the equivalent L-moments derived from the likelihood methods and transfer them to the standard pooling procedures. However, it is unclear what record length should be assigned to the case station, since the effective record length would be somewhere between the length of the systematic record and the length of the period of time in which systematic and historical information are available. Some results for a case study on this simple approach are given below, but more research is needed to provide optimal practical guidelines.

Rather than combining different estimation approaches, it might be possible to use more complex maximum likelihood models. This would allow the pooling together of information from different stations in a unique estimation procedure. Some interesting perspectives on this approach can be found in Nguyen et al. (2014) and in Sabourin and Renard (2015).

An important conceptual difference between practical implementation of the L-moment and the maximum likelihood is that sample L-moments can be estimated for any sample and the decision of which distribution best represents the data at hand can be taken at a later stage of the estimation procedure. In the maximum likelihood approach, however, an assumption on what distribution best fits the data needs to be made as an initial step. Although it is generally assumed that most catchments in the UK follow a GLO distribution, this assumption can be modified if a different distribution seems to be more appropriate for the data under study. Typically, an L-moment diagram can give some indication on whether the sample L-moment seems to be typical for a different distribution than the GLO. In the maximum likelihood approach, some goodness of fitness measures and graphical tools can be used to investigate whether the data seem to give a good fit to the assumed model. However, conceptually maximum likelihood estimation can only be performed for specific possible distributions, which need to be specified before the estimation begins.

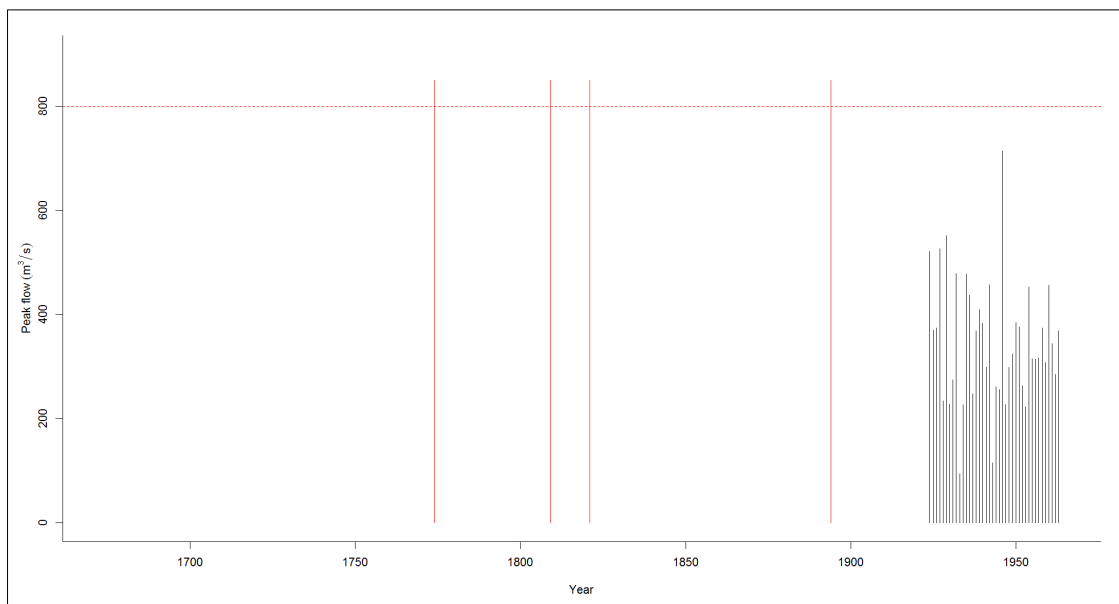
One additional difference between the results obtained in the direct likelihood methods and the pooling strategy outlined in Kjeldsen et al. (2008) is that, while the parameter

estimates for the pooled analysis are constrained so that the growth curve has a value of 1 for  $T = 2$ , no such constrain is placed in the maximum likelihood estimation. This means that the estimated 50th percentile of the distribution might have a different value from QMED. The effect of these constrains in the pooling estimation procedure is not likely to be the cause of large differences in the parameters estimates. It is also worth pointing out that the maximum likelihood approach is directly applied to the original data and not on the sample standardised by the index flood (QMED); the 2 estimation procedures differ in this respect as well.

### B.8.1 Case study

The River Thames at Kingston was chosen as a case study to investigate the effect of combining historical and systematic data into a pooled analysis. Although there is a very long systematic record at this location, only a subset of the available data is used in this case study, namely the data corresponding to the water years between 1924 and 1963. This leaves 40 years of data in the record, which is more representative of the average length of the British records. The specific subset was chosen because it gives at-site estimates similar to those obtained when using the whole series.

The historical information available for the catchment is that 4 events exceeded the perception threshold of  $800\text{m}^3/\text{s}$  between 1673 and 1924. No reliable information could be gathered on the actual values of the flow peak for these historical events, but it is possible to use the information that the perception threshold was exceeded in the estimation procedure. Figure B.38 shows the data used in this case study: for the periods prior to 1924 only the censored information is available.



**Figure B.38 Annual maxima peak flow values used in the case study**

Notes: Censored historical data is shown in red.

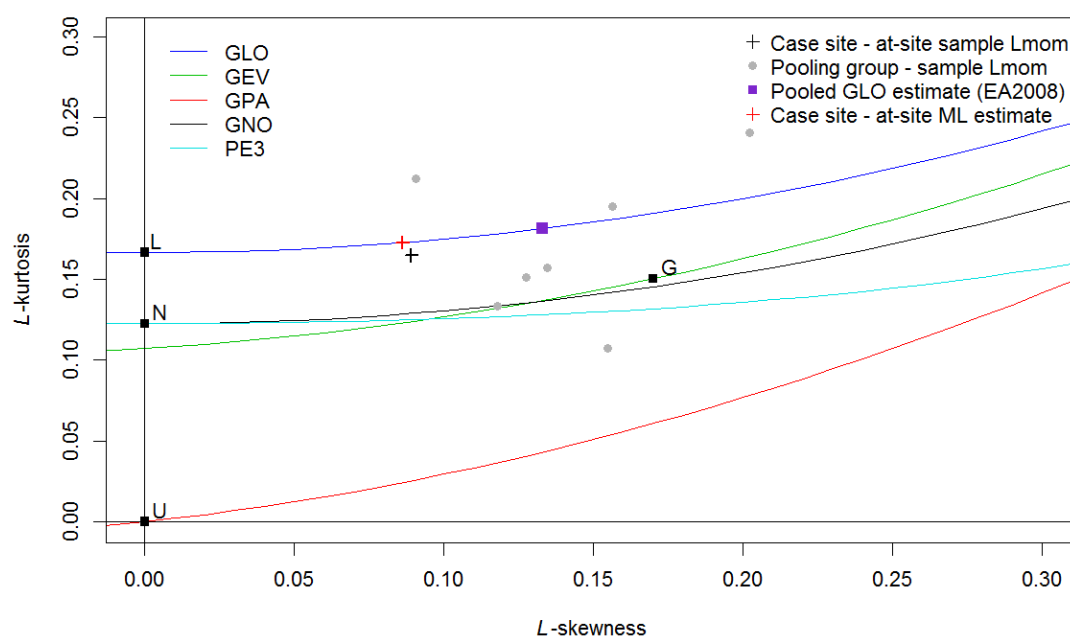
If only the systematic data were available, the estimation procedure according to Environment Agency (2008) would be to:

1. Create a pooling group of catchments similar to the case site.
2. Estimate the sample L-moment from the records in the pooling group.
3. Estimate the sample L-moments for the case site as a weighted average of the pooling group L-moments with weights given according to record length and catchment similarity.

This means that higher weights are given to longer records and to samples of site more similar to the case site, with the case site being given the highest weight.

The estimated sample L-moments can be used to estimate a growth curve of a given distribution by the known relationships between sample L-moments and parameter estimates.

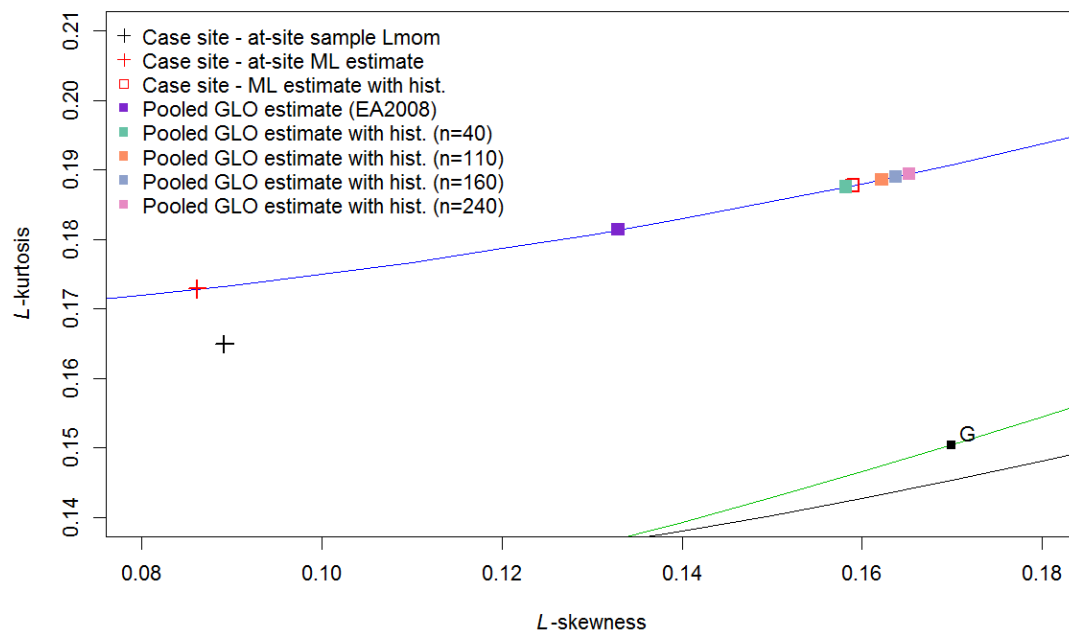
Figure B.39 shows an L-moment diagram with the sample L-moments for the systematic data at the Thames at Kingston (black cross); those of the pooling group are shown as grey dots. The case site is characterised by one of the lowest L-skewness values in the pooling group. The L-moment corresponding to the parameters estimated via the maximum likelihood approach is also shown (red cross). These values are exactly on the GLO line as they were estimated assuming that a GLO distribution was an adequate representation of the distribution of the data for the Thames at Kingston. In fact, the kurtosis is not estimated in the maximum likelihood estimation; instead its value is derived from the constant relationship to the skewness parameter. Assuming a GLO distribution, the pooled parameter can also be estimated and is shown as a purple square in Figure B.39. The pooled estimates have a higher L-skew than the at-site estimate because the records of the pooling group are characterised by L-skewness values higher than that of the case site.



**Figure B.39 L-moment diagram for systematic data only showing at-site and pooled estimates**

Given that historical data are available for the case site, an estimate for the parameters of a GLO distribution could be obtained by combining historical and systematic data via the maximum likelihood method. L-moment estimates corresponding to the GLO parameters obtained via maximum likelihood are obtained by determining the L-moments that would give the same estimates for a GLO distribution as found by the maximum likelihood estimation. The corresponding L-moments are shown as an empty red square in Figure B.40. The estimated parameters correspond to a much higher L-skewness; including historical data in the at-site estimation procedure gives fairly different estimates compared with the systematic only record.

The L-moments corresponding to the estimated parameters using historical data are incorporated into the estimation of pooled L-moments. Initially a record length corresponding to the systematic record length (40 years) is assigned to the case site. Then the record length is gradually increased up to 240 years, the period of time covered by the historical record. The pooled estimates resulting from this exercise are shown in different colours in Figure B.40. The pooled estimates are heavily influenced by the much higher L-skewness obtained when using the historical information for the case site and, in this instance, are very close to the at-site estimates obtained when using historical data. The increasing assigned record length has little effect on the final estimates due to the extremely high value of the L-skewness resulting from using the historical information. Overall, using the available historical information results in much higher estimates for the skewness in the data – ones that are even higher than the skewness estimated using a pooled estimation procedure.



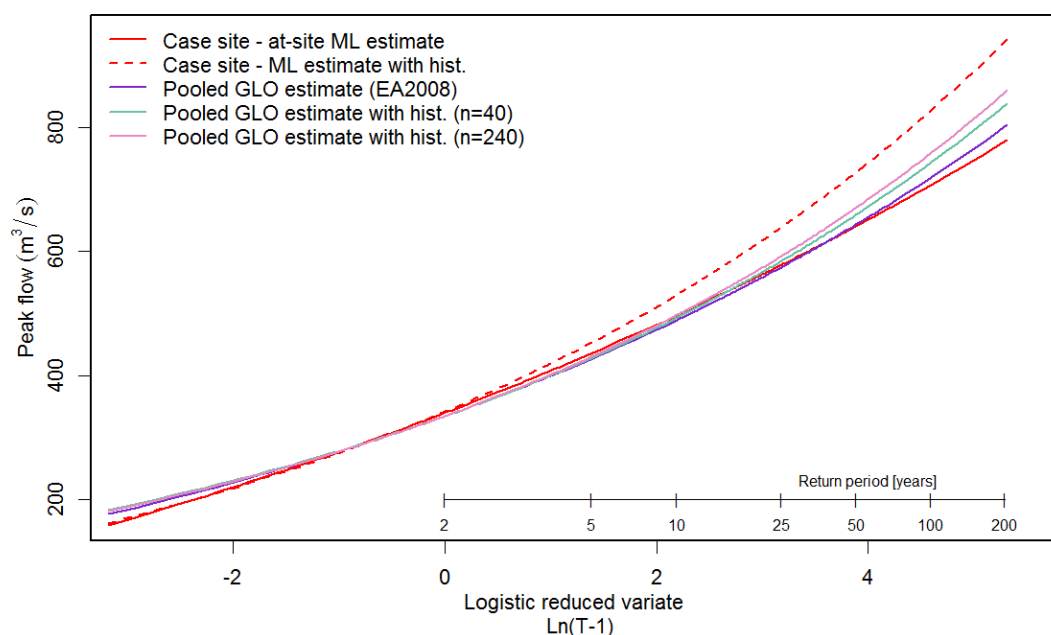
**Figure B.40 L-moment diagram when including historical data: at-site and pooled estimates are shown.**

However, including historical data has a smaller impact on the pooled estimate of the coefficient of variation (CV). Table B.1 shows the estimated L-CV values for the different estimation procedures discussed above. For the maximum likelihood estimation, the reported value corresponds to the equivalent L-CV as calculated via the known relationships between parameter estimates and L-moments. The pooled estimate of the L-CV is smaller than the at-site estimates, and even when giving higher weights to the estimates obtained, including the historical information does not change the final estimates very much.

**Table B.1 Estimated L-CV from different procedures**

| At-site estimates   | L-CV  | Pooled estimates                                   | L-CV  |
|---|-------|--|-------|
| At-site sample L-CV   | 0.192 | Pooled estimate (Environment Agency 2008)          | 0.180 |
| At-site maximum likelihood estimate (equivalent L-CV)                               | 0.190 | Pooled estimate with historical data ( $n = 40$ )  | 0.180 |
| At-site maximum likelihood estimate incorporating historical data (equivalent L-CV) | 0.208 | Pooled estimate with historical data ( $n = 110$ ) | 0.183 |
|   |       | Pooled estimate with historical data ( $n = 160$ ) | 0.183 |
|   |       | Pooled estimate with historical data ( $n = 240$ ) | 0.184 |

Finally, Figure B.41 shows the flood frequency curves estimated via the different methods. The curve estimated using the at-site data with historical information results in higher design flows for long return periods. Including the L-CV resulting from historical information in the pooling procedure leads to steeper flood frequency curves, although the final estimated design events for high return periods would be smaller than those obtained using only the at-site data with historical information.



**Figure B.41 Estimated flood frequency curves according to different methods**

The findings of this case study cannot necessarily be generalised to other records in which systematic and historical data are available. This is because the effect of pooling and including the historical information on the parameter estimates might be very different in different case studies. However, the analysis provides some indications of the possible sensitivity analysis that could be carried out when historical data are available. As pointed out in Bayliss and Reed (2001), the flood frequencies estimates obtained using the historical information should be compared with the estimates obtained using at-site data only and a pooled estimate.

## B.9 Discussion and further perspective

An extensive simulation study has been used to assess the potential benefits of including information of historical peak flow values in flood frequency analysis. The study compared 2 different estimation procedures (maximum likelihood and PPWM) and investigated the effect that mis-specification in the model might have on the quality of the estimation.

Overall, there seems to be convincing evidence that, compared with (P)PWMs, the maximum likelihood approach ensures a larger reduction in the uncertainty of the estimated parameters and design events. Nevertheless, the maximisation of the likelihood can give some numerical problems, while the (P)PWM approach is likely to fail only in very rare cases. The latter could therefore be used as a possible second option or as a tool to give initial values for the optimisation procedure.

The magnitude of reduction in uncertainty in the estimation of design events can be weakened by including some incomplete or imprecise information on the historical records. Although the results from the simulation studies in Section B.7 indicate that the method can be fairly resilient to small mis-specification, in some cases the inclusion of wrongly quantified measures can have a large impact on the overall estimation and give estimates that are biased. If this was the case, the perceived reduction in uncertainty of the estimated design event would actually be detrimental for the understanding of flood magnitudes in a location. When including historical data in a flood frequency analysis, all quantities characterising the historical record should be assessed carefully and some sensitivity analysis to changes in characteristics of the historical record is recommended (see, for example, the case study for the Wear at Durham in the practitioner guide).

The methods illustrated in this appendix can be used to perform an at-site analysis for a specific site for which both historical and systematic peak flow values can be retrieved. Ideally, there would be interest in combining information from sites that might be in the same region but not directly comparable. By extending the maximum likelihood approach, it may be possible to build models to obtain regional estimates in which the historical information of any site can be included in the analysis (this is outside of the scope of this project). A possible avenue of how to combine results obtained using historical and systematic data within a FEH pooled analysis is explored in Section B.8, but further research is needed to provide general guidelines on this application.





# Appendix C: Feasibility study for the development of a Local and Historical Flood Data Archive

## C.1 EXECUTIVE SUMMARY

The FEH Local Project aims to develop new and improved methods and user guidance to better incorporate local, historical and palaeoflood data into Flood Estimation Handbook (FEH) techniques. Although the development of a new system to deliver these data to practitioners is outside the scope of the current proposal, it is envisaged that a future implementation would have greatest benefit if linked to the National River Flow Archive (NRFA), which already holds the core peak flow data used for flood estimation in the UK. A feasibility study was therefore conducted for the development of such a system and its integration with the NRFA. The potential for integration with the Chronology of British Hydrological Events (CBHE) was also explored.

A future implementation of the new system and its supporting procedures would need to have functionality to collate, store, quality control and provide access to the data with a UK-wide remit. The host organisation and project team would need to consider options for the content and technical structure of the system and how it might be set up, operated and maintained. Business planning considerations would include the hosting, funding and governance of the system and the mitigation of risks associated with its use. This feasibility study outlines a number of options in each of these areas and makes recommendations as to the most beneficial. Seven key features (A to G) that would require decisions in a future implementation of the system, are summarised in Table C.1.

In further considering the potential for integration of the new system with the NRFA (Feature A), it is ventured that full integration into the NRFA would be the most effective option, both with regard to costs and to engagement, exploiting existing skills and technical infrastructure, and allowing practitioners to discover and analyse the available information efficiently. In relation to the CBHE, there is consensus that a partnership approach would offer the best opportunity for developing and populating the new system, with the vital engagement of the hydrological community encouraged through the British Hydrological Society (BHS). The recommendation is therefore that **the NRFA remit is extended to host the new system, with the option of integrating all relevant records from the CBHE.**

With regard to database content (Feature B), there is a very broad range of data types that could be of use to FEH practitioners in validating and refining their estimates of peak flows and flood frequency at their site of interest. These include existing estimates of peak flows and related information, hydrometric measurements additional to those already available on national archives, and catchment information such as local amendments to FEH catchment descriptors and changes in hydrological response over time. This feasibility study proposes that hydrometric measurements are best hosted by existing national archives and that catchment information which does not pertain to a specific event are also outside of the scope of the new system. The recommendation is therefore that **the system contains estimates of flood events, quantitative records of flood extents and levels, and the raw observational information such as photographs from which they are derived.**

The technical structure of the system would need to comprise a front-end user interface and a back-end database. The database may be stored either locally, with higher initial costs but greater in-house control, or by a large volume cloud service provider, with long-term costs likely to be higher. The recommended content may include quantitative records of flood level or extent (structured data) and raw (unstructured) records such as a photographs and reports. The recommendation for the back-end database (Feature C) is that it is **held on local storage with in-house systems support, and contains both structured and unstructured data with future phased**

**expansions to accommodate growth in size.** The front-end user interface (Feature D) should be capable of data upload, exploration, display and download and would need to provide different grades of user access. The recommendation is for **an advanced user interface that allows federated searches across different databases, with options for data display and export, and tools for authorised users to add and amend records.**

In setting up the system (Feature E), funding would be required first to establish the systems and procedures and then to populate the empty database with a core set of data. Whilst it would be possible to postpone the population until after the launch, it is recommended that **the system is established along with supporting policies and procedures, and is populated with a core set of data by a consortium of organisations** in order to maximise initial user engagement. The consortium should include: the main Hydrometric Measuring Authorities, that is, the Environment Agency, Natural Resources Wales (NRW), Rivers Agency (RA) and Scottish Environment Protection Agency (SEPA), who have local knowledge of available information, BHS as hosts of the CBHE, CEH as host of the NRFA and leading academics and consultancies in the field who could help with the initial data population.

Post-launch, secure and sustained funding would be required for the ongoing operation and maintenance of the system. Processes would be required for data collation, quality control, data stewardship, system maintenance and user engagement and support. Possible models range from a central hub through which data is collated, checked, improved and extended, to a fully crowd-sourced and crowd-regulated approach in which users can flag and resolve any issues with the data. For data collation and quality control (Feature F) it is recommended that **users submit data, which is quality controlled by central or federated teams of experts, who may also be engaged in proactive collation of data.** For ongoing stewardship of the data (Feature G) it is recommended that **database content is improved by the processing of user data quality flags and by rolling review, and may also be extended by an active programme of extraction of structured data from unstructured records.** A federated approach would provide good access to local knowledge on data quality and relevance.

**Table C.1 Recommendations for the implementation of the new system**

| Feature  | Recommendation  |
|--|---|
| A<br>Integration with existing UK national flood data archives | The remit of the NRFA is extended to host the new system, with the option of integrating all relevant records from the CBHE.  |
| B<br>System content  | The system contains estimates of flood events, quantitative records of flood extents and levels, and the raw observational information such as photographs from which they are derived. |
| C<br>Back-end database   | The system is held on local storage with in-house systems support, and contains both structured and unstructured data. Future phased expansions accommodate growth in size.             |
| D  | An advanced user interface allows federated searches across different (or integrated) databases, with options for   |

|   |  |
|---|--|
| Front-end user interface                | data display and export, and tools for authorised users to add and amend records.  |
| E<br>Setting up the system              | The system is established along with supporting policies and procedures, and is populated with a core set of data by a consortium of organisations.  |
| F<br>Data collation and quality control | Users submit data, which is quality controlled by central or federated teams of staff, who may also be engaged in proactive collation of data.   |
| G<br>Data stewardship                   | Database content is improved by the processing of user data quality flags and by rolling review, and may also be extended by an active programme of extraction of structured data from unstructured records. |

Business planning matters include hosting, funding, governance and risk mitigation. The system would be best hosted by a single organisation with secure long-term status that is able to publish data in an open and accessible way for the UK as a whole. Funding commitment would be required both to the initial development, and importantly, to the ongoing costs of operating and maintaining the database. Ideally this would be secured on a five or ten year, or even ongoing basis as the risks of proceeding without long-term funding include wasted resources, deterioration of good practice in flood estimation techniques and reputational damage to the host, partners and funders. A Project Board would be required to oversee the initial development of the system. It is recommended that ongoing governance is provided by the UK Surface and Groundwater Archives (SAGA) Committee. A technical group comprising representatives of the developers, users and Measuring Authorities would facilitate future developments.

In conclusion, the delivery of a system to meet the aims of the project is eminently feasible, given clearly defined limits on the data types to be included and secure funding, with a national remit, both for the establishment and for the long-term operation and maintenance of the system. The position of the NRFA as the UK's central archive of peak flow data, with proven capability in database and web development and in hydrological data stewardship, and with established partnerships with the primary Measuring Authorities recommends it as a suitable host. There would be significant benefits of integrating the system with existing national peak flow databases (primarily the NRFA and CBHE) and agreement would need to be reached with the organisations hosting these services as to the extent and nature of their involvement and changes required to these existing systems.

## C.2 INTRODUCTION

### C.2.1 Aims

This study addresses Task 4b of the FEH Local Project to develop a “proposed new approach to collect, store, quantify quality and provide access to local, historical and palaeoflood data that can be used in design flood estimation”.

The project brief sets the development of a new system to deliver these data to practitioners outside of the scope of the current proposal; however, it envisages that a future implementation would be linked to the existing NRFA which provides data to support flood estimation in the UK. This document takes the form of a feasibility study for this future implementation, laying out the context and discussing the possible content, structure, development and operation of such a system, and the policy and business planning considerations that would need to be addressed in its delivery.

Embedded within the document are Option Boxes for seven features of the system. These identify key decisions that would be required in delivering a solution, and present a range of options with varying degrees of functionality. The options are discussed and recommendations made in Section C.8 Conclusions and Recommendations.

### C. 2.2 Background

This section provides historical context through a brief history of the development of flood data archives in the UK, and geographical context through some examples of national databases in other countries that make flood data available through a web-based portal.

#### C.2.2.1 Historical UK national flood data archives

A historical overview of initiatives to archive flood event data is provided by the NRFA website (Centre for Ecology & Hydrology, 2016).

In 1975, instantaneous flood peaks for over 550 gauging stations were published in Volume IV for the Flood Studies Report (NERC, 1975) along with tabulated catchment characteristics and flood statistics. A major update to flood peak data, begun at the Institute of Hydrology in April 1985 under Ministry of Agriculture, Fisheries and Food funding, was completed in October 1991 (Bayliss and Jones 1993). This database contained over 77,000 peaks from 859 gauging stations throughout the UK, ending in or before 1990. The FEH (Institute of Hydrology, 1999) updated the flood peak dataset to finish in 1994. It also launched version 1.0 of the WINFAP-FEH program that makes use of the peak flow data to statistically estimate the flood frequency curve at a given location on the river network.

Between 2004 and 2014 UK flood peak data were provided via the HiFlows-UK website, hosted on the Environment Agency website. HiFlows-UK was a partnership between the UK Measuring Authorities – the Environment Agency in England and Wales, SEPA in Scotland and the Rivers Agency in Northern Ireland. In 2001, the project was awarded a grant from HM Treasury's Capital Modernisation Fund. The HiFlows-UK data were an updated and enhanced version of the data contained in Volume 3 of the FEH, increasing the data available by 40%. HiFlows-UK contained significantly more data than the FEH dataset, incorporating the results of subsequent data reviews to improve data quality, and additional information on each station,

including the indicative suitability of the data. The data and website was initially released as a pilot for testing and feedback on both the website functionality and data in March 2004. The first full version of the WINFAP-FEH data (v1.1) were released on 1 August 2005 and the last version released under the HiFlows-UK project (v3.3.2) was released in April 2014. JBA Consulting carried out a number of updates to the HiFlows-UK database between 2004 and 2014 and facilitated the v2.1.1, v3.02, v3.1.2 and v3.3.2 releases.

In April 2014, responsibility for the provision of UK national flood peak data was transferred to the NRFA, maintained by CEH in close collaboration with the UK hydrometric Measuring Authorities. Funding for its future maintenance is being provided jointly by CEH, the Environment Agency, SEPA, the Rivers Agency and NRW. Governance of the dataset has been merged with that for other national hydrometric data by placing it under the remit of the UK SAGA Committee comprising UK and devolved Government, the Natural Environment Research Council (NERC), CEH, British Geological Survey (BGS), Environment Agency, SEPA, Rivers Agency, NRW, Met Office, Canal and Rivers Trust, BHS, Chartered Institute of Water and Environmental Management and the water industry.

### **C.2.2.2 Examples of historical and local flood data archives in other countries**

*“The use of geospatial databases for the visualisation of information and capability to embed images within such databases presents an important development, permitting flood levels and additional information beyond a basic descriptive account to be housed within each flood account, empowering the researcher to more rapidly and easily access required information. One of the principal constraints to the wider application of historical information in flood frequency analysis has been the time requirements for collecting the necessary data; well developed and constructed geospatial databases present a valuable step towards removing these constraints.”* Kjeldsen et al. (2014)

Examples of national flood data archives of the type applauded by Kjeldsen *et al.* (2014) have been found in a number of other countries, with various purposes, intended audience, data types, technical solutions and host organisations.

In Australia, the Australian Flood Risk Portal (Geoscience Australia, 2016) aims to provide public discovery, visualisation and retrieval of flood studies, flood maps, satellite derived water observations and other related information. It is hosted by Geoscience Australia, a government funded organisation, and includes a web-based portal that allows registered users to submit flood study reports, and open access for the search and display of information via an interactive map.

In Ireland, a National Flood Hazard Mapping Website has been developed by the Office of Public Works (2016). This aims to provide information to planners and the public to identify areas at risk of flooding. Content includes reports, photographs, newspaper articles and other information about reported floods, and is searchable by location including an interactive map.

The French La Base de Données Historiques sur les Inondations (Historical Database on Floods) (IRSTEA, 2016), has search and explore operations in its web-based user interface and user registration for government departments and institutions involved in the management of flood risk. There is open access to historical data from as early as the 13<sup>th</sup> century which is plotted alongside recent flood occurrence, and details of the damage caused by recent events may be viewed and downloaded free of charge.

Although regional in geographical scope, the closest example to the current proposal found is the Colorado Flood Database (Colorado Water Science Centre, 2016; Kohn, 2013) which aims to collate and provide access to historical flood data with the aim of

improving flood frequency estimation by engineers, scientists and water resource managers. The database is hosted and regularly updated by the Colorado Water Science Centre, of the U.S. Geological Survey (USGS), and is government funded. Data include indirect discharge measurements stored in USGS offices, floods from indirect discharge measurements referenced in USGS reports, palaeoflood studies from peer-reviewed journal articles, and the USGS National Water Information System peak-streamflow database. It is populated centrally, and its web-based user interface includes data exploration, display and export functions and hyperlinks to source material.

### **C.2.3 Existing UK National Flood Data Archives**

It is assumed that the new system would take the form of an electronic database, implemented either as an extension to an existing archive, or entirely independently. This section describes existing national flood data archives in the UK and considers the potential for integration.

#### **C2.3.1 Current NRFA**

The NRFA is the primary archive of daily and peak river flows for the United Kingdom. Its staff collate, quality control and archive hydrometric data from gauging station networks across the country including the extensive networks operated by the Environment Agency, NRW, SEPA and the Rivers Agency. Daily, monthly and flood peak data are held from over 1400 gauging stations and the archive incorporates a wide range of hydrological information to assist in the understanding and interpretation of measured river flows.

The flood peak data is available for a subset of 958 stations and comprises Annual Maxima (AMAX) and Peaks Over Threshold (POT) data, ratings and datums as well as FEH catchment descriptors. AMAX, POT and catchment descriptors can be downloaded from the website for use in the WINFAP-FEH software. The new FEH Web Service, also hosted by CEH, provides an interactive map-based web portal for catchment descriptors and rainfall data.

Further details of the NRFA can be found in Dixon et al. (2013) and on its website at <http://nrfa.ceh.ac.uk>.

#### **C.2.3.2 BHS Chronology of British Hydrological Events**

The Chronology of British Hydrological Events (CBHE) was launched in 1998, with the aim of improving access to historical hydrological information. It takes the form of a database of text records with an on-line user interface allowing searches of existing records and the addition of new records by users (Law *et al.*, 2016). The database is not limited to flood events or river flows, but is open to any information pertaining to rainfall or runoff.

Each record is required to include specific information regarding a hydrological event, preferably using a source quotation, a source reference, the date of occurrence and a geographical reference. In the first four years of operation, over 7300 records, requiring 10 MB of server space, were contributed to the database, principally drawn from periodicals known to contain detailed accounts of hydrological events (Black and Law, 2004).

The University of Dundee has hosted the CBHE from its inception, supporting its development with limited resources. In June 2016, after a period offline, the website

was relaunched, hosted by BHS reflecting its recognition in the hydrological community. User registration has been revised such that login is required to contribute to the database, but searching of its contents is unrestricted. The relaunch includes capability for geo-referencing of records that can then be displayed on an interactive map.

Further details of the CBHE can be found on its website at <http://www.cbhe.hydrology.org.uk>.

### **C2.3.3 Potential for integration of the new system with existing archives**

Initial discussions on the potential for integration of the new system with the CBHE took place in February 2016 between representatives of the NRFA and BHS. There was broad consensus that, should the development of the new system proceed, a partnership approach would offer the best opportunity for its development and population, with the vital engagement of the hydrological community encouraged through the BHS. Ideas were raised surrounding the degree of integration of the databases and these are summarised in Option Box A.

Option A1, in which the new system is developed as a standalone system would not require the involvement of the NRFA or the CBHE and therefore would not benefit from the tools and procedures already established or the skills and experience of the associated staff. The usefulness of such a system to the practitioner would be limited as it would be unlikely to deliver efficiencies, becoming an additional source of information to mine with considerable potential for duplication in existing archives.

Option A2, in which the systems are linked but remain largely separate, would offer time efficiencies in search operations, but no advantage over A1 in data exploration or download. The partnership of the NRFA, the CBHE and BHS in such an approach could deliver some sharing of staff skills, tools and procedures given adequate and sustained resources and clear governance.

Option A3, in which the new system is populated with data from the CBHE and fully integrated into the NRFA, would benefit from the tools, procedures, skills and contacts pooled from the NRFA, the CBHE and BHS. Sustained extension to the resources of the NRFA would be necessary so that sufficient staff time could be allocated to data stewardship and user support. Similarly, additional developers would be required to implement and maintain the systems and introduce new operations such as user authentication. However, since the skills required to carry out these tasks are tried and tested at the NRFA, and its reputation as a provider of high quality data is established, the costs and risks may be lower than those associated with an independent approach.

Further consideration is given to the potential advantages of integration in the sections on the technical structure (Section C.4) and population (Sections C.3.4 and C.5.2) of the system.



#### **Option Box A – Integration with existing UK national flood data archives**

- A1. Existing UK national flood data archives databases (CBHE and NRFA) remain unchanged, new system developed as an entirely separate entity.
- A2. Records in an adapted CBHE and the new system cross-referenced to one another, both capable of interaction with the NRFA.
- A3. NRFA remit extended to host the new system, with the option of integrating all relevant records from the CBHE.

### **C.2.4 Geographical and historical scope**

Although the current project is funded through England's Environment Agency, it is strongly recommended that the new system should be UK-wide in geographical scope. Failure to collate the data nationally would create a two-tier capability for flood estimation in the UK, for example meaning that an engineer might benefit from local validation of their design not available to their counterparts in Scotland, Wales or Northern Ireland. An England only system would also prevent the use of local and historical information in the assessment of instrumental records from other parts of the UK which are being utilised in FEH pooling groups for English sites. Establishing a UK wide system need not require multiplication of resources if the database is hosted by an national organisation such as CEH and managed with consistent UK wide policies, procedures and governance (as is currently the case with the NRFA).

It is further recommended that in designing a national system, consideration is given to maximising its interoperability with other potential international systems which might be developed in future, such as the European system advocated by Kjeldsen *et al.* (2014).

With regard to historical scope, it is proposed that there be no restrictions on the period covered by the database, other than those imposed by copyright law (see Sections C.5.1 and C.7.4). This is because records that could be of use in flood estimation may exist for current, historical or even palaeoflood events.

### **C.2.5 Stakeholders**

The intended users are FEH practitioners who may include consultants, regulators, local authorities and construction engineers. Interest is also likely from the wider hydrological and scientific community and potentially from a very broad audience including local communities, the media, local government, the insurance industry and others with an awareness of historical or current flooding. Additional stakeholders include potential funders (Section C.7.2) and the operators of the existing national flood archives such as CEH and BHS.

## C.3 DATABASE CONTENT

### C.3.1. Types of local and historical flood data

**Table C.2 Data groups**

| Data group   | Data types   |
|--|--|
| Additional hydrometric measurements                        | <p>River flow and stage records from established gauging stations already forms the basis for flood estimation in the UK. The core hydrometric information used for such purposes is collated, quality controlled and made available through the NRFA. However, at a local scale additional hydrometric records which are not currently held in a central archive are potentially useful. Such information has, until now, not been deemed useful enough to include in the national archives and includes:</p> <ul style="list-style-type: none"> <li>• Flow records from gauging stations of uncertain accuracy</li> <li>• Flow records from temporary gauging stations</li> <li>• Level only records from gauging stations</li> <li>• Groundwater level records, further to the 181 sites on the National Groundwater Level Archive (NGLA)</li> <li>• Spot measurements of river flow or level (at gauging stations or otherwise)</li> </ul> |
| Additional catchment information                           | <p>Catchment data including FEH catchment descriptors are already made available by CEH. A wide range of additional information can be used to improve the practitioner's understanding of local catchment hydrology. Such information may be newly generated for the project or obtained from a wider range of potential sources. It may take the form of additional new information or local-scale alterations to information that is already published. This includes:</p> <ul style="list-style-type: none"> <li>• Local-scale channel characteristics (e.g. width)</li> <li>• Local amendments to FEH catchment descriptors</li> <li>• Amended catchment boundaries, especially for small catchments</li> <li>• Hydrological and hydraulic processes involved in flood generation</li> </ul>  |
| Estimates of ungauged flood events and related information | <p>For flood events that were not measured by hydrometric instrumentation at the time of occurrence, other observations can be used to generate an estimate of the event magnitude that can be ranked relative to other events. In conducting future estimations of flood risk, practitioners may want to</p>  |

|  |   |
|--|---|
|  | <p>reuse the estimate of the flood event and/or the observational information. This category of information includes:</p> <p>a) Estimates of individual flood events (flow/time/location)</p> <p>b) Observational information that has been or could be used to generate/refine the flow estimate of such events. Such information usually pertains to an observation of the level and/or extent of a flood and includes:</p> <ul style="list-style-type: none"> <li>• Photographs of floods</li> <li>• Details of physical flood marks</li> <li>• Palaeoflood records</li> <li>• Documentary records of a flood (e.g. reports, papers, media, local history and anecdotal records, weather diaries)</li> <li>• Number of flooded properties and other impacts</li> </ul> |
|--|---|

The priority data types for inclusion in the new system should necessarily be those of most value in current industry standard flood estimation techniques and in the methodologies proposed in this project. There is a wide variety of data types referenced in the practitioner guidance document *Using local data to reduce uncertainty in flood frequency estimation*, prepared by JBA Consulting for Task 3a of the project. These fall into three broad groups; hydrometric measurements, catchment information and records and estimates of flood events (Table C.2).

### C.3.2 Recommended data scope

Whilst it would be possible to design a system capable of handling all three types of data, the most appropriate solutions for collating and providing access to each data type at a national scale are likely to differ. Each of the different groups of data is very different in its technical nature, source, structure, ownership and likely use. This section outlines the key differences and suggested development routes.

#### C.3.2.1 Additional hydrometric information

Well established national archives already exist in the UK for the collation, stewardship, storage and retrieval of river flow data (NRFA) and groundwater level data (NGLA). At present, these archives are selective in the information that they hold and do not aim to store all available hydrometric information. For example, the NRFA only contains river flow records from gauging stations that the operating Measuring Authority and NRFA consider capable of producing data of an acceptable standard.

Stations where records are considered of poor quality, short-term temporary stations and stations recording river stage but not calculating flow are not stored on the NRFA or NGLA. However, stage data are very valuable in flood estimation, and can be rated back to flow estimates in future or reconciled with hydraulic models. The additional hydrometric information considered in this data group therefore relates to such stations and also includes spot gauging of river flow at gauging stations or other points on the river network. In terms of their technical content, data structure and sources, such information is however very similar to that currently held on the NRFA.

This report therefore considers that rather than set up a new separate source of additional hydrometric information, such data would be best held as an extension to the current national archives. The NRFA and NGLA provide suitable platforms for holding temporary and spot records, and measurements of uncertain quality, and while additional funding would be needed to facilitate such extensions and the ongoing data stewardship, this presents the most logical and cost-effective solution for this data group.

This report will not provide further details on the options and feasibility of storing data in this group and it is recommended that separate discussions be initiated between the hydrometric Measuring Authorities and the NRFA/NGLA regarding options for progressing such archive extensions. Such an initiative would be consistent with the UK Government's move towards Open Data (Defra, 2016) and could encourage a national, collaborative approach, seeking to maximise the effective use of resources and the quality of the data provided.

### **C.3.2.2 Additional local catchment Information**

Local catchment information derived by a practitioner, such as measurements of channel width and hydrological processes which do not pertain to a specific event, are also outside of the scope of the new system. No database can replace the experience and knowledge of the local practitioner, nor should it discourage the good practice of visiting sites.

Information on developments affecting the temporal or spatial variability of hydrological response, such as flood schemes, impoundments and land use changes such as urbanisation, provide important context for event data. Further discussion is required on how this may best be made available; perhaps through future development of the existing system that provides catchment information to users, within the wider context of improving the handling of non-stationarity in UK flood estimation.

### **C.3.2.3 Estimates of ungauged flood events and related information**

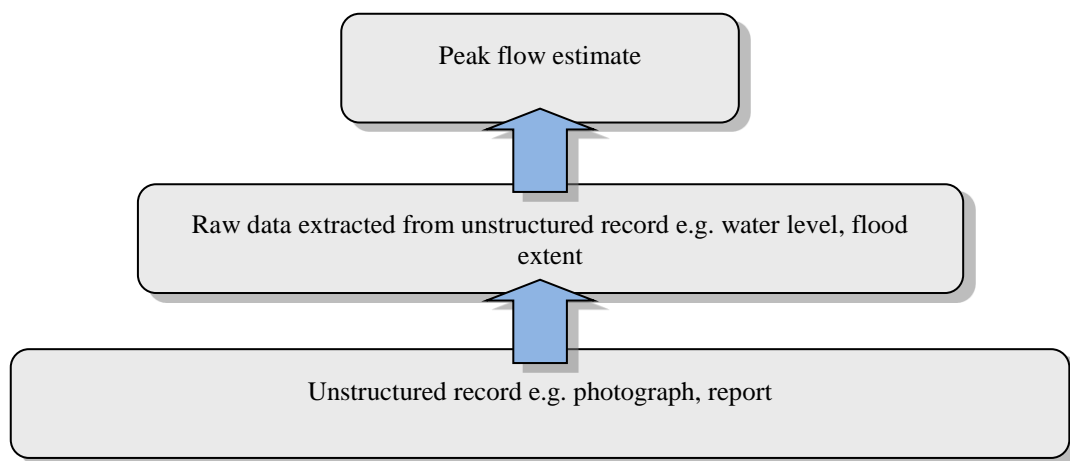
Records of flood events that were not measured by hydrometric monitoring systems and the observational information that has, or can be used to derive or refine such estimates represents the key information that it would be practical and useful to collate but for which a suitable national data archive does not currently exist. As such, the remainder of this report will focus on this data group.

## **C.3.3 Hierarchy of data to be included**

Within the **estimates of ungauged flood events and related information** group are three levels of data (Figure C.1) with usefulness to FEH practitioners dependent upon the amount of preparation required before the data is in a useable form. The least ready level is an unstructured record such as a photograph or report. Many, but not all, of these, will contain some information that can be extracted into a quantitative observation of flood level or extent. These in turn, may be used in statistical methods of flood frequency and peak flow estimation.

Whilst extraction would require some processing of observational records, the focus of the system content should be on the factual recording of events. Users would be encouraged to use a description field to record any qualification or pertinent details relating to location or circumstances at the time of an observation that could assist the practitioner in the task of interpretation. Under the hierarchy shown, details of

assessments that have been carried out, such as flood events used and resulting growth curves, would be held as unstructured data or referenced with a link, although future developments could consider holding such information in structured form (see Section C.4.2.1).



**Figure C.1 Hierarchy of data usefulness**

Option Box B presents three approaches to database content, which correspond to the three levels in the hierarchy of data usefulness shown in figure C.1. Whilst interest to other users is also a consideration, this is secondary to the needs of the intended audience. The technical considerations in handling the various data types are also of fundamental importance, and these are addressed, and options presented, in Section C.4.

#### **Option Box B – System content**

- B1. Estimates of flood events (flow/stage/date\_time/location) with citation or link to a source report explaining how it was generated and text description
- B2. Estimates of flood events (flow/stage/date\_time/location) with citation or link to a source report explaining how it was generated and text description and quantitative information about flood extents/levels with source of observation
- B3. Estimates of flood events (flow/stage/date\_time/location) with citation or link to a source report explaining how it was generated and text description and quantitative information about flood extents/levels with source of observation and raw observational information (copies/links to media reports, photographs, etc.)

### **C.3.4 Potential data sources**

A wide range of data sources exists, falling into three broad categories. The first is existing systems, such as the CBHE, from which a large number of text records may be suitable for populating the new system (Section C.5.2). The second is existing data that are not so readily accessed, but may be extracted from consultancy flood studies,

academic literature and historical reports. The third comprises sources of data that have yet to be recorded, which may include physical flood marks, newspaper reports, local community knowledge and social media. Suitable methods of data collation will depend upon the data source in question (Section C.6.1).

### **C.3.5 Data quality**

Three elements of data quality would need to be considered:

- i) Errors – such as incorrect data types, spurious and missing data;
- ii) Uncertainty – encompassing both precision and accuracy; for example a stone floodmark on a bridge is precise, but is of uncertain accuracy, a newspaper report describing a flood extent is likely to be accurate in identifying which areas of a town were affected, but imprecise;
- iii) Usefulness – an estimate of a flood peak for a historical or palaeoflood event may have high uncertainty, but if it is the only record of that event, it may nevertheless be very valuable in reducing uncertainty in flood estimation.

A system of classification would be required to facilitate the user in understanding the quality of the data and its appropriate use. Processes for the quality control of submitted data and for the ongoing stewardship of database content would need to consider which of these can be addressed (Section C.6).

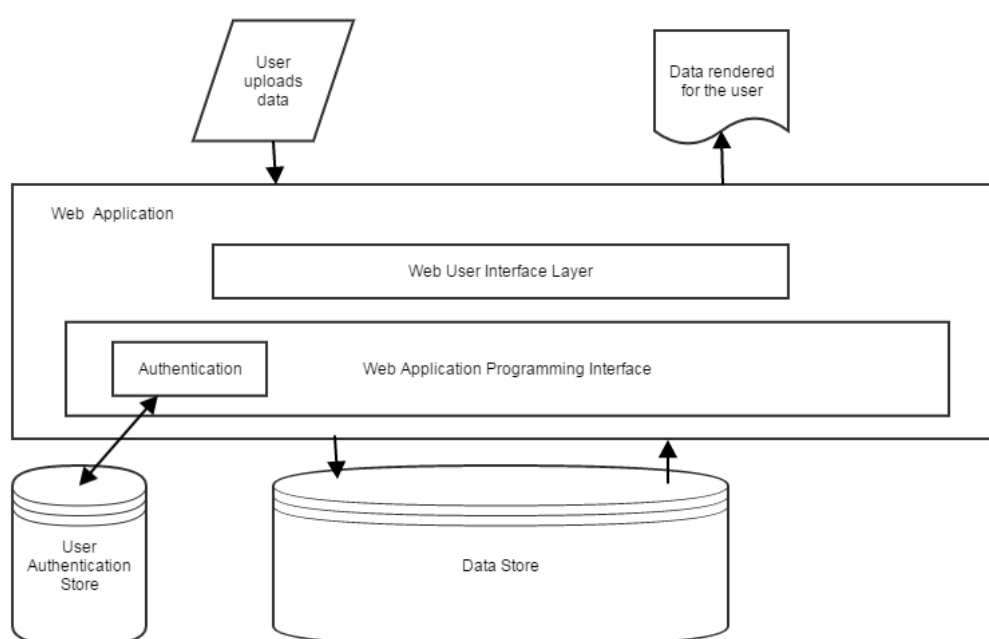
## C.4 TECHNICAL STRUCTURE

The technical requirements of a system capable of handling the data described in Section C.3.3 and providing an interface for their users are outlined in this section.

### C.4.1 System overview

The core of the system would be the data store, the back-end database. A web application built on top of the data store would provide interface for all types of including the administration (figure C.2) **Figure C.2 Conceptual scheme of the system and interaction with users**

).



**Figure C.2 Conceptual scheme of the system and interaction with users**

### C.4.2 Back-end database

Relational databases such as Oracle, which is used by CEH to host the NRFA, offer the functionality and flexibility required for the storage of the data. If well designed, these allow the data to be stored, searched, linked and viewed in many different ways without requiring restructuring.

#### C.4.2.1 Structure of records

For the purposes of data handling, there are two basic types of data: structured and unstructured. Structured data are organised into fields (columns) within the database making it possible to sort, link, query and analyse the data. Examples include

estimates of peak flow, with an associated location and date/time, and measurements of water level, with an associated location, date/time and local datum.

Unstructured data are those that cannot be organised into fields, and include photographs, videos and reports. They could also include links to data and reports that are available elsewhere in the public domain, such as a flood chronology study for a given location and time period.

All items submitted to the database would require:

- i) A location (grid reference, lat/long, town name, river name, area etc.)
- ii) A date/time (a single instant or a range, with varying accuracy and resolution)

This would be the minimum requirement to make the data searchable in a relational database. Most other datatypes would also require one or more text fields that could be indexed to assist with searching. In addition to content, all data types would have associated metadata such as the date of entry, the source of the material and a description.

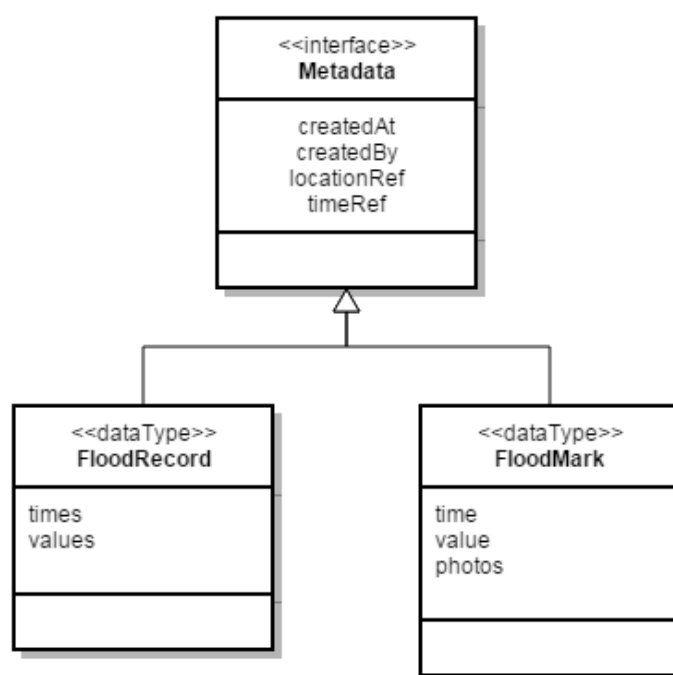
#### **C.4.2.2 Database structure**

Since the new system may be required to store a wide variety of data types, from structured records of location and magnitude to unstructured collections of photos, graphs, information about flood marks and newspaper articles, an unstructured data store is required.

In its simplest form, this would hold data of all declared types along with basic metadata for each record. To allow specifications of different data types, specialised data stores could be created, based on relational databases (figure C.3). These specialised data stores would allow streamlined access and analysis of the data optimised for the particular data type.

Data would need to be supplied in agreed units, scales, reference systems, coded values, and other domains. It is essential to agree and define these domains for each data type before any data of that type are inserted into the system and that any inserted data are automatically validated using these domains.





**Figure C.3 Illustrative example of specialisation of data types**

There would be marked differences between the data as seen by the user, and the underlying data structure. For example, a system of data stores would be required to track the quality control processes. From a use point of view, this may appear as a simple but searchable flag showing its status as “unchecked”, “quality control in progress” or “quality controlled”.

#### **C.4.2.3 Data storage considerations**

The primary considerations relating to data storage are location, size and reliability and each has implications for cost. The database may be stored locally or on the cloud, with size highly dependent on the inclusion of images, and with backup arrangements that should be selected to secure the desired degree of reliability.

If stored locally, the database would need to be administered by a host organisation who would procure, maintain and update the hardware. This would require a considerable initial outlay, then steady or phased increases in funding to meet the costs of ongoing general computer administration and upgrades to the hardware. This is the model used for the current NRFA. By contrast cloud storage is a service, for which a provider charges an annual fee for a given amount of storage, for example, 100 GB for £25 per year. Since there are no hardware costs, initial outlay is lower, but costs would grow at a faster rate than those of local storage as they are more sensitive to the size of the database. In the long-term, therefore, cloud storage would be more expensive, but is a lower risk option for any pilot project. This decision would depend on the expected lifetime and size of the database.

The potential size of a database is greatly increased by the inclusion of images which are a desirable and versatile source of information. The current NRFA requires 8.7 Gb of storage for the data alone, and 15.7 Gb with the associated images. A single, historical photograph is around 1 Mb, a modern, uncompressed, high resolution

photograph 20 Mb and videos are at least an order of magnitude larger. Hence the inclusion of images on local storage would require controls on their number, size and format.

The reliability of a database is expressed in terms of percentage uptime, and costs are incurred by duplicate backups and systems support. For example, the “five nines” standard for high availability means that a system is available 99.999% of the time, allowing downtime of less than 6 minutes a year and requiring a backup system that can be brought online instantaneously. It is anticipated that the required reliability of the new system would be similar to that of the current NRFA, which is supported in-house during office hours from CEH with a backup duplicate system that can be brought online within a few working days.

Three options for the database are presented in Option Box C. The first two of these require local storage; C1 for a simple system with only unstructured data stores, C2 for a system storing both structured and unstructured data. Option C3 outsources the data storage to a cloud service provider, with fewer concomitant constraints on size, but less control over systems and future costs. The benefits of integrating the system with the NRFA include in-house expertise in relational databases, and for local storage options, in-house control over uptime.

#### **Option Box C – Back-end database**

- C1. Local storage with in-house systems support and content limited to unstructured data below a given size. Future phased expansions to accommodate growth in size.
- C2. Local storage with in-house systems support, content both structured and unstructured data, with size restrictions on submissions. Future phased expansions to accommodate growth in size.
- C3. Large volume, cloud-based storage, capable of storing both structured and unstructured data.

### **C.4.3 Front-end user interface**

Users would interact with the system through the Web User Interface Layer (figure C.2) – the “front-end”. In line with the capabilities of the back-end system, the front-end should allow the following operations:

- Discover (search items by specified parameters, can be considered bulk display operation);
- Display (display details in several forms);
- Download (individual and bulk);
- Create new item(s);
- Edit or update existing item(s).

The discover and display operations are expanded in Section C.4.3.2 and the create and download operations in Sections C.4.3.1 and C.4.3.3 respectively. The ongoing editing of data is covered later in Section C.6.2 on Data Stewardship. The key pages (or views) that would need to be developed in an initial version of the website are:

- Home (landing page) – main entry point to the whole system;
- About – page with detailed information about the system, help, etc.;
- Upload item – form to upload new data;
- List items – view of data available in the system with mechanisms to select items by parameters;
- Item details – detailed view of an item;
- User profile – user specific page with information relevant to the user.

The bulk of the interface visible to the user would be implemented in HTML, CCS, and JavaScript.

#### **C.4.3.1 Data input**

The user would be presented with a form and prompted to select data to upload from their hard drive. In an initial version, the system could rely on precise and potentially extensive user input (such as selecting options in many dropdown boxes, entering text, etc.) with automatic input validation performed server side. Over time, the system would aspire to extract properties from the files selected by the user and assist the user by offering options relevant only to the data being uploaded.

If initial data submission were by email, data would be inserted through the website interface by central staff. Operations that would be tedious to do or repeat using the website interface can be automated using the application programming interface and many scripting languages (Python, R, bash, JavaScript, etc.).

#### **C.4.3.2 Data discovery and display**

The system must allow users to search for content and offer relevant results. It is often good practice to offer a simple search interface and an advanced search interface. While the results of the simple search and advanced search are rendered the same way (e.g. a list of items, a map with item locations displayed, etc.), the advanced search allows users to fine tune the parameters of the query. A few examples below illustrate the kinds of search parameters users may want to specify:

- Display all items (there would likely be a maximum limit of items to display and a paging mechanism);
- Display only curated items;
- Display only items uploaded in, say, 2019;
- Display only flow records where flow is higher than 1 cubic metre per second.

The system should be able to display the search results in several forms. For example:

- A plain list of records similar to a search engine results page;
- A table;
- Items on an interactive map;
- In a graph where applicable.

It is important to consider options for displaying information within the context of other data that reside in a different system. For example, the interactive map could load extra layers of information from Web Map Services published by CEH or other organisations, such as the location of gauging stations. Similarly, the graph view could display daily flow or peak flow data published by other organisations as web services enabling the user, for example, to view historical and instrumental period data alongside each other. The ability to discover and display data from various sources together offers opportunities to capitalise on the efforts invested into these systems.

#### **C.4.3.3 Data download**

Items can be rendered either into the browser window for display or to the user for download. For example, a photograph of a historic flood mark can be rendered as a point on a map with the photograph attached and/or offered for download in a zip archive with all metadata included. A bookmarking mechanism would allow users to add items to their user profile page from which items can be revisited at a later date.

If the new system were to be integrated with the NRFA, there would be scope to link data download operations as well as the discover and display operations. For example, a download of the WINFAP-FEH files from the peak flow database could include any historical flood peaks that exist at, or near to, the location of interest. The new system would also benefit from existing engagement of the user community with the NRFA; the peak flow pages of the NRFA website, which include a download link for WINFAP-FEH data, had 8,443 visits between late March and December 2014.

#### **C.4.3.4 Controlling user access**

Implementing user authentication is necessary in order to allow upload and/or stewardship of content through the web interface. Different users would be able to complete different tasks depending on their role. At least four basic roles would be considered (table C.3).

**Table C.3 User roles**

| <b>Role</b>     | <b>Description</b>  |
|-----------------|---|
| Unauthenticated | Members of the general public and visitors who do not have a user account or who are not logged into the system. As a minimum, these users would be able to see the homepage, information on how to sign up, the overall description of the system and the data or functions it offers, and information about policies. |
| Public          | Members with basic level of access, able to access information and data that has been cleared for public access. These users should be able to upload content and data.   |
| Private         | Could do everything that Public users could do, and more. This role would probably need breaking down into more roles. For example, users would be able to upload content and keep it visible only to them. Other users may be able to review and approve or disapprove content or other users etc.                     |

|               |   |
|---------------|---|
| Administrator | Administrators who would have access to the administrative interface. |
|---------------|---|

The NRFA currently has no user authentication because data is submitted and uploaded centrally by CEH and there is no upload facility in the user interface. The new system, whether or not it is integrated with the NRFA, would require appropriate privacy policies to be in place for the secure input and storage of user information, such as an email addresses and encrypted passwords.

The choices pertaining to the front-end are myriad and Option Box D presents a somewhat simplistic set of scenarios which would need to be developed in consultation with stakeholders. In practice, operations would be driven by the data held, the degree of integration with existing databases and the user requirements and would be likely to evolve as the new system became established.

#### **Option Box D – Front-end user interface**

- D1. Basic data search and display system. No web-based data upload functionality.
- D2. Advanced search and display capabilities. Several user levels and ability to upload content for selected users.
- D3. Advanced user interface allowing federated searches across different (or integrated) databases, options for data display and export, and authorised users to add and amend records.

## **C.4.4 Maintenance tools**

### **C.4.4.1 Administration interface**

The administrative interface would be part of the same website but access to administrative tasks would be available to authorised users only. Most of the administrative tasks would be performed through the website and the web application would include mechanisms for tracking changes.

All tasks, including administrative operations, could be performed using the website. However, some one-off complex and bulk adjustments, imports, or exports may be too costly to implement in the web application and direct access to the data stores may be more feasible.

The user interface would rely on an application programming interface (API), which would provide programmatic access to the system and therefore allow more custom applications to be developed.

In addition to the views mentioned above, at least the following administrator views would be required:

- Update Item – page to edit properties of an item

### **C.4.4.2 Data validation and editing tools**

If the system contains only unstructured data (Option C1), no validation or editing would be required. For structured data (Options C2, C3 or similar), a suite of tools would be required to facilitate quality control procedures and editing. For example, the upload operation could be supported by an error checking tool that compares data identified by the user as flow measurement against limits held by the system and asks for confirmation in a pop up window if any values are outside of these limits.

Protection against malicious uploads such as computer viruses would need to be embedded into any upload tools. Identification of false data that a user may wish to upload in order to skew a flood estimation and justify a higher flood barrier is more challenging and would require quality control by hydrologists.

Quality control procedures would require a plotting tool to show the data in graphical form, and a spatial display tool for exploration of the records within a catchment. These would be made available to authorised users and could assist in the identification of duplicate, malicious or erroneous entries. Flagging and editing tools would be required to identify suspect entries and make corrections to the data or metadata submitted. The tools developed by CEH for use with the NRFA for such purposes include a Time Series Plotter (TSP) and a Metadata Editor.

#### **C.4.5 Hardware considerations**

Requirements for hardware can be established only when number of user requests per hour and similar metrics are known or estimated. Initial versions would target industry standard hardware with a configuration equivalent to virtual machines available on cloud hosting services.

Advancements in network connectivity mean that the physical location of the hardware is not important. For example, the cloud-based option for data storage (Option C3) would probably mean that the hardware supporting the front-end and back-end systems would be in different places.

## C.5 SETTING UP THE SYSTEM

This section considers the two development phases that would need to be planned and resourced prior to the launch of the new system; firstly establishment of the system and procedures, and secondly population with an initial set of data.

### C.5.1 Establishing the system and procedures

A project to deliver the technical systems described in Section C.4 would need to extend to the procedures and guidance governing and encouraging its use. These would need to include the establishment of tools and procedures for data collation, quality control and stewardship (Section C.6), determination of data policies, training of staff and plans for community engagement (Section C.6.5.1).

Policies and user guidance would be required to cover citation of sources, intellectual property and restrictions on use, and to mitigate the risks of malicious upload and breaches of personal data protection. Such policies must be communicated to users via clearly displayed terms and conditions, user declarations and disclaimers. For example, the NRFA has policies on its website covering the ownership of the data and a liability disclaimer concerning its quality.

Copyright infringement is prevented by the CBHE by accepting source material no more recent than 70 years before the present, and requiring the consent of authors or publishers for bulk use (> 5%) of a single source. This is conditional upon the purpose of the copying being research or private study (Black and Law, 2004). Copyright law allows historical information within such constraints, and more recent material if unrestricted. The referencing of source material as links or citations, rather than holding copies on the system, may be a suitable alternative (see Option Box B).

### C.5.2 Populating with data

A decision would be required on whether the project to set up the system should extend to its population with an initial set of data (Option Box E). An initial population could include adding flood event observations from the CBHE (Option A3), ingesting tables of historical flood flows and levels in Volume 4 of the Flood Studies Report (NERC, 1975) and actively sourcing and adding some of the existing local and historical data held by the research community. This would be best delivered as a consortium involving key data holders in the initial set up project. Such an approach would encourage uptake of the system, once launched, as there would already be some useful data available. At this stage, submission of data by email might be practicable, with uploading to the database conducted by administrative users. The consortium should include Measuring Authorities, BHS members, CEH, JBA Consulting and leading academics in the field such as Dr Andrew Black of the University of Dundee and Dr Neil MacDonald of the University of Liverpool.

**Option Box E – Setting up the system**

- E1. Empty system is created according to options selected for back-end and front-end, policies and procedures are established and staff trained.
- E2. Empty system is created (E1) and populated with a core set of data by a partnership of organisations.



## C.6 OPERATION AND MAINTAINANCE

This section considers the processes and procedures that would be required to operate and maintain the system after its launch. These are concerned with the continued collation, quality control and stewardship of data, maintenance of the systems and user engagement and support.

### C.6.1 Data collation

Once established, the ongoing input of data to the system by all users would be via the web-based user-interface, encouraged through techniques of user engagement (Section C.6.5.1) and where appropriate, regulation.

The collation of information from the local community, for example, such as photographs of physical flood marks on buildings and bridges, could be gathered using “crowd-sourcing” techniques, whereby a large number of people are encouraged to engage with the system through the web-based user-interface, contributing their knowledge and observations. The success and longevity of a crowd-sourced approach would be highly dependent on engagement with the hydrological and local communities.

In the case of consultancy reports, since many of the projects making use of the database are likely to be publicly funded, it is anticipated that the principle of open data could be upheld, perhaps through a contractual stipulation that flood estimates refined with data from the system are themselves submitted to the system, reinforcing and enhancing its utility as a new national database. In this way, a partly regulated approach may be used to increase the amount of data held.

An alternative approach to data collation is to establish a project for the ongoing, proactive mining of the data from the academic literature, consultancy reports, local communities and other sources of information described in Section C.3.4. Such a project would benefit from the involvement of historians and archivists as well as hydrologists.

### C.6.2 Quality control

User authentication would need to identify errors in the submitted data so that spurious data could be removed. It might also extend to hydrological assessment of the precision and accuracy of the data. This may be done centrally, by the Measuring Authorities or other federated group of organisations or volunteers, or by the crowd. Numerous possible models of quality control therefore exist:

- 1) Central quality control of all data by the system custodians;
- 2) Light touch quality control by central custodians, checking data formats/completeness but not the hydrology;
- 3) Federated quality control by funded regional Measuring Authorities;
- 4) Federated quality control by a team of volunteers;
- 5) Open access WIKI approach with quality control conducted by the crowd;
- 6) No quality control.

The approach taken by the Australian Flood Risk Portal is for nominated reviewers to conduct checks for accuracy and completeness, with national oversight provided by the

custodians, Geoscience Australia. Users finding incorrect data are invited to contact the custodians, and thus contribute to ongoing data stewardship (Section 6.3). A similar model that uses reviewers with hydrological expertise is desirable in mitigating the potential risk of circulation and reinforcement of errors in the data if a regulated approach is taken to collation.

A selection of three possible approaches to data collation and quality control are presented in Option Box F. The NRFA currently collects data by email in an annual submission, with quality control shared with the Measuring Authorities. Such an approach is likely to be unsuitable for the new system because the amount of data likely to be submitted is unknown and the timing of submission cannot be coordinated in the same way. The assumption of user input is therefore made. The federated involvement of the Measuring Authorities would bring valuable local knowledge on data quality and relevance through their area and regional staff.

#### **Option Box F – Data collation and quality control**

- F1. Users input data directly to the system, data undergoes no or crowd-based quality control.
- F2. Users submit all data to the system, data undergoes quality control by central or federated teams of staff.
- F3. Users submit data to the system with regulated involvement from publicly funded projects; data is quality controlled by central or federated teams of archive staff who are also engaged in proactively seeking data.

### **C.6.3 Data stewardship**

The ongoing maintenance of content should allow for identification of issues, editing of records and extraction of structured from unstructured data.

#### **C.6.3.1 Data improvements**

Whilst data could be accepted without further editing once on the system (see Option Box G), the establishment of formal management and challenge processes may be considered more appropriate for the intended application in decision making and flood risk management. Ongoing improvement of the data could include metadata editing, linking of records that relate to the same event, assessment of contradictory values and removal of duplicates and erroneous data. A system of flags through which the custodians and/or users could identify suspect, incomplete or edited data, or newly recovered or interpreted data (Bayliss and Reed, 2001) would be an advantage. A rolling review programme, such as the ongoing review of the peak flow data at the NRFA, would ensure that the content represented the best available collation of historical and local flood data in the UK. Similarly, an auditable query process, like the one used for updates to the NRFA, would seek to ensure that consistent decisions are made on whether to uphold or reject changes.

The NRFA is well placed to deliver such operations as staff with hydrological expertise and data and validation tools (Section C.4.4.2) are required.

### **C.6.3.2 Extraction**

Further enhancement of the content could be delivered by the extraction of structured data from unstructured records, that is, movement from the bottom to the middle of the hierarchy of data presented in figure C.1. For example, on uploading a photograph, the user could be prompted to identify that it contains a measurement for future extraction, and the variable measured. Failure to include extraction of this type as an integral part of the ongoing data stewardship may affect the usefulness of records, and ultimately the likely level of engagement with the database.

#### **Option Box G – Data stewardship**

- G1. No editing once data is on the system.
- G2. Database content improved through the processing of user flags and rolling data review.
- G3. Database content improved by the processing of user flags and rolling data review and also extended by an active programme of extraction of structured data from unstructured records.

### **C.6.4 System maintenance**

The ongoing funding required for systems maintenance would be dependent upon the decisions made with regard to data storage (Section C.4.2.3 and Option Box C) and the front-end user interface (Section C.4.3 and Option Box D).

## **C6.5 User engagement and support**

### **C.6.5.1 User engagement**

A communication and outreach strategy would need to be established to engage both the hydrological and local communities, so as to maximise the number of data users and contributors and mitigate the risks associated with a lack of engagement (Section C.7.4).

As discussed in Section C,2.3.3 on integration, the engagement of the hydrological community, during both the setting up and ongoing operation of the system, would best be delivered in partnership with BHS. This could include presentations at meetings, journal articles, social media and a launch event to raise awareness and explain the purpose and potential value of the system.

Plans for engagement with local communities should draw on the expertise in citizen science used for the gathering of ecological information at CEH and include outreach of the sort already conducted by NRFA staff through their Twitter account and newsletter. The same methods of communication could be used to keep users up to date with developments on the database and draw attention to useful additions and data types.

A group of stakeholders requiring special consideration would be members of the press, who might be interested in using photographs, videos or flood peaks to prepare articles and reports during a flood event. It would be wise for the host organisation to be one with experience in handling press enquiries of this nature.

### **C.6.5.2 Hydrological support**

A user support helpdesk would be required, of the sort provided by the NRFA to support the current holdings of daily, monthly, peak flow and catchment data. This would answer enquiries about the data contained in the system, log reports of potential data issues/errors and answer general questions related to the system. The helpdesk would need to be staffed by hydrologists with expertise in hydrometry and FEH techniques.

While the system would be focused on data provision, rather than methodologies for its use, another of the risks associated with the database is that information might be incorrectly applied in FEH techniques. This is mitigated by the practitioner guidelines *Using local data to reduce uncertainty in flood frequency estimation*, prepared by JBA Consulting. Furthermore, since the existence of the database would serve to encourage and inform the use of validation in flood estimation techniques, the uncertainties and associated risks of harm to property or people should be reduced. Nevertheless, the policies established in setting up the system (Section C.5.1) would need to cover the eventuality of a warranty claim concerning its accuracy, and liability for any damages.

Any future editions of the practitioner guidelines should be prepared in consultation with the database custodians so that the techniques of validation described are supported by the stored data as far as is appropriate. Similarly, future developments of the database should be done in conjunction with the custodians of the guidelines so that they continue to meet the aims of the project.

### **C.6.5.3 Systems support**

User guidance for interaction with the database through the web user interface would need to be integral part of the website. System support would also need to be provided through the user helpdesk (see Section C.6.5.2). A feedback form should be made available as part of the website, to allow users to report issues and request new features at any time. Feedback provided should be periodically reviewed by staff, dealing with urgent issues on an operational basis and filing long-term issues into a development backlog for the next update. Behind the scenes, documentation for administrators and application programmers would be required as offline documents.

## C.7 BUSINESS PLANNING

The primary concern in business planning of the database is sustainability. Each decision regarding the hosting, funding, governance and mitigation of risks should be made with a view to the long-term viability of the system.

### C.7.1 Hosting

Assuming a national remit is preferred, the database would be best hosted by a single organisation with secure long-term status that is able to publish data in an open and accessible way for the UK as a whole. The organisation should have capability and proven staff skills in database and web development, hydrological data stewardship and user engagement. However, a consortium approach to the population of the new system is advocated (Section C.8.2), and the long-term involvement of partner organisations would be required in a federated approach to quality control (Section C.6.2).

### C.7.2 Funding

The development of a UK-wide system would require the involvement of funders in addition to the Environment Agency, such as SEPA, NRW and the Rivers Agency, who are broadly supportive of the project's aims, although not providing financial support at this stage. Other funding sources could be explored but it is likely that financing for the set up and ongoing maintenance of such a system would need to come from the public purse and the nature and benefits of the data are such that a pay-per-access or subscription based model is unlikely to be appropriate. Furthermore the success of any system would be dependent on the community contributing content in a free and open manner.

A commitment would be required both to the initial development, and importantly, to the ongoing costs of operating and maintaining the database. Ideally this would be secured for a five or ten year, or even ongoing basis as the risks of proceeding without long-term funding include wasted resources, deterioration of good practice in flood estimation techniques and reputational damage to the host, partners and funders.

### C.7.3 Governance

A Project Board would be required to oversee the initial development of the system. There would be benefits from ongoing governance by the UK SAGA Committee. A technical group comprising representatives of the developers, users and Measuring Authorities would facilitate future developments.

### C.7.4 Risks and mitigation

Risks that might arise from the development and operation of the new system are summarised in table C.4, with some suggestions for mitigation measures as described in earlier sections of the report.

**Table C.4 Risks and mitigation**

| <b>Hazard</b>   | <b>Harm</b>                                       | <b>Severity</b> | <b>Likelihood</b> | <b>Mitigation measures</b>   |
|---|---|-----------------|-------------------|--|
| Poor flood estimates arising from poor data or methodology    | Damage to property and potential risk to life     | High            | Medium            | Quality control of data (Section C.6.2)<br><br>Practitioner guidelines and user support for flood estimates (Section C.6.5.2)<br><br>Warranty and liability policies (Section C.5.1) |
| Insufficient regard for data ownership by developers or users | Breaking the law                                  | Medium          | High              | Policies on intellectual property, personal data protection, open data and restrictions on use. Minimum workable amount of personal data held. (Section C.5.1)                       |
| Malicious uploads   | Damage to the IT systems of the host organisation | Medium          | Low               | Technical resilience and citation of sources<br><br>User declaration (Sections C.4.4.2 and C.5.1)  |
| Poor longevity  | Wasted resources and reputational damage          | Medium          | Medium            | Maintenance of systems and content secured by ongoing funding and delivered by a robust organisation (Sections C.7.1 and C.7.2)  |
| Poor engagement   | Wasted resources, “white elephant” system         | Medium          | Medium            | Hosting, access and outreach designed to maximise engagement (Section C.6.5.1)   |
| Misuse of information   | Reputational damage                               | Medium          | Medium            | Press engagement (Section C.6.5.1)   |

## C.8 CONCLUSIONS AND RECOMMENDATIONS

### C.8.1 Conclusions

The FEH Local Project aims to develop new and improved methods and user guidance to better incorporate local, historical and palaeoflood data into FEH techniques. This report takes the form of a feasibility study for the development of a new system that along with supporting procedures, has the functionality to collect, store, quality control and provide access to these data.

The delivery of such a system is eminently possible, with five important caveats:

- 1) The types of data selected for inclusion – and conversely, outside of the scope of the system – must be clearly defined;
- 2) The geographical scope should be UK-wide, with the support of funders and regulators in England, Scotland, Wales and Northern Ireland;
- 3) Funding must be secured both for the set up and for the long-term operation and maintenance of the system;
- 4) The host organisation must have capability and proven staff skills in database and web development and hydrological data stewardship and be prepared to work in partnership with other organisations to facilitate population of, and user engagement with the database;
- 5) There would be significant benefits of integrating the system with existing national peak flow databases (primarily the NRFA and CBHE); agreement would need to be reached with the organisations hosting these services as to the extent and nature of their involvement and changes required to these existing systems.

### C.8.2 Recommendations

Seven Option Boxes throughout the report provide a focus for some of the key decisions that would need to be made in the implementation of the new system. These are reproduced in table C.5, with recommendations and supporting justification.

**Table C.5 Option recommendations**

| Box   | Options  | Recommendation and justification   |
|---|--|--|
| A:<br>Integration with existing UK national flood data archives | A1. Existing UK National Flood Data Archives databases (CBHE and NRFA) remain unchanged, new system developed as an entirely separate entity.<br><br>A2. Records in an adapted CBHE and the new system cross-referenced to one another, both capable of interaction with the NRFA. | <b>Recommendation: Option A3</b><br><b>NRFA remit is extended to host the new system, with the option of integrating all relevant records from the CBHE.</b><br><br>Further discussion with partnership organisations is required, but the recommendation from the NRFA is that if a commitment is to be made to meeting the aims of the project, full integration of the systems would be |

|                         |   |  |
|-------------------------|---|--|
|                         | A3. NRFA remit is extended to host the new system, with the option of integrating all relevant records from the CBHE.   | the most cost effective option. This is likely to be the most efficient both in exploiting existing skills and technical infrastructure, and from the practitioners' perspective in discovering and analysing the available information.   |
| B:<br>System content    | <p>B1. Estimates of flood events (flow/stage/date_time/location) with citation or link to a source report explaining how it was generated and text description</p> <p>B2. In addition to Option B1, quantitative information about flood extents/levels with source of observation</p> <p>B3. In addition to Options B1 and B2, raw observational information (copies/links to media reports, photographs, etc.)</p>  | <p><b>Recommendation: Option B3</b></p> <p><b>The system contains estimates of flood events, quantitative records of flood extents and levels, and the raw observational information such as photographs from which they are derived.</b></p> <p>Restricting the content of the system to the most readily usable (Option B1) or quantitative (Option B2) data would result in extremely useful information being lost to the practitioners just because it is in photographic form. Whilst requiring funding for a more complex system, this option carries less risk of creating a "white elephant" through lack of engagement.</p>  |
| C:<br>Back-end database | <p>C1. Local storage with in-house systems support and content limited to unstructured data below a given size. Future phased expansions to accommodate growth in size.</p> <p>C2. Local storage with in-house systems support, content both structured and unstructured data, with size restrictions on submissions. Future phased expansions to accommodate growth in size.</p> <p>C3. Large volume, cloud-based storage, capable of storing both structured and unstructured data.</p> | <p><b>Recommendation: Option C2</b></p> <p><b>The database is held on local storage with in-house systems support, and contains both structured and unstructured data. Future phased expansions accommodate growth in size.</b></p> <p>The 'quick and dirty' solution (Option C1), whilst simple from a systems point of view, is not capable of storing the highest priority data in its most readily usable form (as required for each of the options in Box B). Whilst Option C3 is best placed to meet modern expectations of data storage derived from social media, the higher long-term costs reduce its appeal. Option C2, including some restrictions on the size of uploads, represents the best value for money, especially when hosted by an organisation with existing local storage capability and procedures.</p> |
| D:                      | D1. Basic data search and display system. No web-based data upload functionality.   | <p><b>Recommendation: Option D3</b></p> <p><b>An advanced user interface allows federated searches across different (or integrated) databases, with</b></p>  |



|  |  |   |
|--|--|---|
| Front-end user interface                 | <p>D2. Advanced search and display capabilities. Several user levels and ability to upload content for selected users.</p> <p>D3. Advanced user interface allowing federated searches across different (or integrated) databases, options for data display and export, and authorised users to add and amend records.</p>                        | <p><b>options for data display and export, and tools for authorised users to add and amend records.</b></p> <p>The email submission of data implied in Option D1 is unsuitable for ongoing data collation (see Option Box F below) and Option D2 is therefore the minimum that is workable in practice. Whilst development of the front-end would be ongoing, Option D3 is recommended as a starting point so that search and display operations allow historical and recent peak flows to be explored together.</p>  |
| E:<br>Setting up the system              | <p>E1. Empty system is created according to options selected for back-end and front-end, policies and procedures are established and staff trained.</p> <p>E2. Empty system is created (Option E1) and populated with a core set of data by a partnership of organisations</p>   | <p><b>Recommendation: Option E2</b></p> <p><b>The system is established along with supporting policies and procedures, and is populated with a core set of data by a consortium of organisations.</b></p> <p>It is anticipated that engagement with the system would be greatly enhanced by launching it ready-populated with an initial set of data. The consortium should include Measuring Authorities, BHS members, CEH, JBA Consulting and leading academics in the field such as Dr Andrew Black of the University of Dundee and Dr Neil MacDonald of the University of Liverpool.</p>  |
| F:<br>Data collation and quality control | <p>F1. Users input data directly to the system, data undergoes no or crowd-based quality control.</p> <p>F2. Users submit all data to the system, data undergoes quality control by central or federated teams of staff.</p> <p>F3. Users submit data to the system that is quality controlled and data is also proactively sought by staff.</p> | <p><b>Recommendation: Option F2 or F3</b></p> <p><b>Users submit data, which is quality controlled by central or federated teams of staff, who may also be engaged in proactive collation of data.</b></p> <p>Formal management of options F2 and F3 would deliver the standard expected of NRFA data and required for its application in flood risk management. Option F2 would require a regulated approach, whereby publically funded projects concerned with flood risk assessment were contractually obliged to upload their data to the system. Option F3, building on techniques used in populating the CBHE, would require sustained funding of data collation staff and provide the opportunity to</p> |

|                           |  |  |
|---------------------------|--|--|
|                           |  | steer its content in order to maximise the usefulness of the system.   |
| G:<br>Data<br>stewardship | <p>G1. No editing once data is on the system.</p> <p>G2. Database content improved through the processing of user flags and rolling data review.</p> <p>G3. Database content improved by the processing of user flags and rolling data review and also extended by an active programme of extraction of structured data from unstructured records.</p> | <p><b>Recommendation: Option G2 or G3</b></p> <p><b>Database content is improved by the processing of user data quality flags and by rolling review, and may also be extended by an active programme of extraction of structured data from unstructured records.</b></p> <p>If taking the regulated approach (Option F2), Option G2 in which the content is improved but not actively extended would suffice, as the onus of extraction could be placed on the user. If data were to be actively sought by staff (Option F3), an accompanying programme of extraction would be an integral part of the work programme for a centrally funded or federated teams.</p> |

### C.8.3 Future developments

The aim of the current project in serving FEH practitioners limits the content to high flow data and supporting evidence. In considering other users, there would be benefit in extending the system to include a wider range of observation hydrological information, as has been the practice of the CBHE. Examples include high rainfall events and low extremes such as dry reaches. Each potential extension would require careful appraisal of the content, technical systems, setting up, operation, maintenance and business plans involved.

## C.9 REFERENCES

- Bayliss, A. C. and Jones, R. C. (1993). Institute of Hydrology Report No 121. ISBN: 0948-540-47-8.
- Bayliss, A.C. and Reed, D.W. (2001). The use of historical data in flood frequency estimation. Report to MAFF. CEH Wallingford.
- Black, A. R. and F. M. Law (2004). Development and utilization of a national web-based chronology of hydrological events. *Hydrological Sciences Journal* 49: 237-246.
- Centre for Ecology & Hydrology (2016). National River Flow Archive. [online] Available at: <http://nrfa.ceh.ac.uk/> [Accessed 1 Mar. 2016].
- Colorado Water Science Centre (2016). Colorado Flood Database. [online] Available at: <http://cwscpublic2.cr.usgs.gov/projects/coflood/COFloodMap.html> [Accessed 1 Mar. 2016].
- Department for Environment, Food & Rural Affairs (2016). Open Data Programme. [online] Available at: <https://data.blog.gov.uk/2015/06/29/defra-announces-major-open-data-programme/> [Accessed 23 June 2016]
- Dixon, H., Hannaford, J., Fry, M. J. (2013). The effective management of national hydrometric data: experiences from the United Kingdom. *Hydrological Sciences Journal* 58(7): 1383-1399.
- Geoscience Australia (2016). Australian Flood Risk Portal. [online] Available at: <http://www.ga.gov.au/flood-study-web> [Accessed 1 Mar. 2016].
- Institute of Hydrology (1999) Flood Estimation Handbook, (five volumes). Institute of Hydrology, Wallingford. ISBN: 978-0-948540-94-3.
- IRSTEA National Science Research Institute and Technology for the Environment and the Agriculture (2016). La Base de Données Historiques sur les Inondations. [online] Available at: <http://bdhi.fr/appli/web/welcome> [Accessed 1 Mar. 2016].
- Kjeldsen, T. R., Macdonald, N., Lang, M., Mediero, L., Albuquerque, T., Bogdanowicz, E., Brazdil, R., Castellarin, A., David, V., Fleig, A., Gul, G. O., Kriauciuniene, J., Kohnova, S., Merz, B., Nicholson, O., Roald, L. A., Salinas, J. L., Sarauskiene, D., Sraj, M., Strupczewski, W., Szolgay, J., Tournazis, A., Vanneuville, W., Veijalainen, N. and Wilson, D. (2014). Documentary evidence of past floods in Europe and their utility in flood frequency estimation. *Journal of Hydrology* 517: 963-973.
- Kohn, M. S., Jarrett, R.D., Krammes, G.S., and Mommandi, Amanullah (2013). Web-based flood database for Colorado, water years 1867 through 2011. U.S. Geological Survey: 26.
- Law, F. M., Black, A.R., Scarrott, R. M. J., Miller, J.B. and Bayliss, A.C. (2016). Chronology of British Hydrological Events. [online] Available at: <http://www.trp.dundee.ac.uk/cbhe/welcome.htm> [Accessed 1 Mar. 2016].
- NERC (1975). Flood Studies Report. (five volumes). Natural Environment Research Council, London. ISBN: 0901875252 9780901875259.
- Office of Public Works (2016). Australian Flood Risk Portal. [online] Available at: <http://www.floodmaps.ie> [Accessed 1 Mar. 2016].

**Would you like to find out more about us  
or about your environment?**

**Then call us on**

**03708 506 506** (Monday to Friday, 8am to 6pm)

**email**

**enquiries@environment-agency.gov.uk**

**or visit our website**

**[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)**

**incident hotline 0800 807060** (24 hours)

**floodline 0345 988 1188 / 0845 988 1188** (24 hours)

Find out about call charges ([www.gov.uk/call-charges](http://www.gov.uk/call-charges))



**Environment first:** Are you viewing this on screen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don't forget to reuse and recycle if possible.